# STUBBORN SYSTEMS:

# TWO NEW MULTIPLE COHERENT SYSTEMS OBJECTIONS FOR COHERENTIST MORAL REALISM

A Thesis

by

ROSS T. COLEBROOK

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF ARTS

August 2011

Major Subject: Philosophy

Stubborn Systems: Two New Multiple Coherent Systems Objections for Coherentist Moral Realism

# STUBBORN SYSTEMS:

# TWO NEW MULTIPLE COHERENT SYSTEMS OBJECTIONS FOR COHERENTIST MORAL REALISM

A Thesis

by

ROSS T. COLEBROOK

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF ARTS

Approved by:

| | |
|---|---|
| Chair of Committee, | Gary Varner |
| Committee Members, | Linda Radzik |
| | Jim Grau |
| Head of Department, | Daniel Conway |

August 2011

Major Subject: Philosophy

# ABSTRACT

Stubborn Systems: Two New Multiple Coherent Systems Objections for Coherentist Moral Realism.

(August 2011)

Ross T. Colebrook, B.A., American University

Chair of Advisory Committee: Dr. Gary Varner

In *Moral Realism and the Foundations of Ethics*, David Brink defends a version of moral realism that makes use of a coherentist epistemology. I argue that we have good reasons to prefer a coherentist epistemology in ethics, and that common arguments in favor of foundationalism and against coherentism fail. I conclude that only the multiple coherent systems (MCS) objection poses a serious threat to coherentism, and that there is a way of meeting this objection as it is traditionally construed. Doing so requires adding an explanatory requirement to coherentism: a coherent system must not only *be* coherent, but explain *why* it is coherent.

Though I argue that the MCS objection fails as traditionally construed, I propose variants of the objection which apply specifically to coherentist moral realism. Based on the social intuitionist model of moral decision-making, I point to evidence which indicates that intuitive moral judgments are the cause of reasoned moral judgments. This raises two problems: "the problem of intransigence" and the "limited problem of intransigence." These two problems constitute MCS objections to coherentist moral

realism. The first problem claims that coherentism itself might not be possible for actual human beings, and that even if people are capable of forming coherent systems of belief, these systems are necessarily radically divergent as a result of various cultural and sociological factors. The second problem is that the coherentist moral realist should not expect a single moral system to result from moral reasoning (even ideally), because all such systems will include intuitions which are immune to reasoning. I argue that these MCS objections must be overcome if coherentist moral realism is to be a viable option.

# TABLE OF CONTENTS

# 1. INTRODUCTION

One of the most frustrating aspects of metaethics is the seemingly inconclusive nature of many of its most prominent debates. Though there is a wealth of literature on the epistemic justification of moral beliefs, there has been far too little in the way of decisive arguments. The reason for this may be the interconnected nature of many metaethical debates: if one believes that moral judgments are fact-stating and truth-functional, it is hard to maintain a dialogue over epistemic justification with someone who believes that moral judgments are merely expressions of emotions which lack truth-functionality.

Another failing of the prominent metaethical debates is the relative dearth of empirical data used to support any given side. Cognitivists and noncognitivists, internalists and externalists, coherentists and foundationalists, all seem content to propose arguments from the armchair. I think these arguments are seldom decisive, not for a lack of imaginative thought experiments and argumentative rigor, but for a lack of hard *facts* about the psychological processes behind the moral decision-making of real people. I believe that psychological research can shed light on many traditional debates in metaethics, and this thesis (particularly section three) attempts to do just that in the case of moral epistemology.

With these two points in mind, I propose that one see this thesis as an attempt to provide a critique of coherentist moral realism which avoids a heavy reliance on (what I see as) stale and non-empirical metaethical debates. I intend to do this by arguing

---

This thesis follows the style of *The MLA Handbook.*

against coherentist moral realism while committing to as many of its main tenets as possible. Thus I treat moral beliefs as cognitive, truth-functional, and aimed at objective facts at the outset, and avoid taking any position whatsoever on such troublesome issues as moral internalism and externalism. This strategy of agreement will, I hope, prove more convincing than one that simply pits traditional opposites against each other.

Nonetheless, in order to give the coherentist moral realist a fair shake, I will need to pay homage to the tradition. My arguments in section two are not intended to be a thorough overview of the debate between coherentists and foundationalists, but I do aim to at least defend coherentism against some of its most prominent challenges. In keeping with this strategy, section two will argue that (1) coherentism need not collapse into constructivism, (2) two of the traditional arguments against moral foundationalism (intuitionism) are quite strong, and (3) three of the most important standard objections to coherentism fail. Even the strongest of these objections (the non-moral multiple coherent systems (MCS) objection) fails, though as you will see, some detailed work on the concept of explanation is necessary to show why this is so.

Where the second section of this thesis is mostly concerned with traditional arguments, the third section is concerned with what I take to be somewhat novel ones, and it is in this section that I draw on experimental evidence about the role of intuitions in moral judgments. The social intuitionist (SI) model of moral decision-making indicates that intuitive moral judgments are not changed by reasoning (they are "intransigent"), and that reasoned moral judgments are always ultimately caused by intuitions. I will refer to this thesis as the "problem of intransigence" and argue that it

constitutes an MCS objection against coherentist moral realism, because it implies that people of different cultural and sociological backgrounds can *never* arrive at convergent systems of moral belief. I will also point to a "limited problem of intransigence," which arises even if we think that reasoned moral judgments are not *all* caused by intuitive moral judgments. This problem holds that we have reason to believe that any coherent moral system will include at least some intransigent intuitions. This constitutes an MCS objection because, if these intransigent intuitions infect all coherent moral systems, coherentism seems sure to lend justification to many different competing systems of morality. These two forms of the MCS objection are stronger than the non-moral MCS objection, and they imply that coherentist moral realism must either turn subjectivist or become immensely implausible. These objections must be overcome if coherentist moral realism is to be a good account of moral justification.

## 2. COHERENTIST MORAL REALISM

In this section I will define and defend coherentism in general (and coherentist moral realism in particular) against a number of prominent critiques in the literature. In subsection section one, I will present David Brink's coherentist moral realism. In subsection section two, I will characterize coherentism, as well as its most well-known alternative, foundationalism. In subsection three, I will argue that coherentism in ethics does not imply constructivism in ethics. In subsection four, I aim to outline the main critiques of foundationalism from a coherentist perspective. In subsection five, I examine three important criticisms of coherentism: the claim that it is viciously circular, the isolation problem, and the multiple coherent systems (MCS) objection. I will argue that all three criticisms fail, but that the MCS objection is the strongest. In subsection six I will elucidate Philip Kitcher's account of explanation as the unification of our beliefs, and in subsection seven I will argue that we can use this account of explanation to overcome the MCS objection.

### 2.1 Preliminary Distinctions

In explaining Brink's position in moral epistemology, I must first distinguish his moral realism from some of its most prominent alternatives. Brink (1989) defines realism in ethics as any position which accepts the following two basic tenets: "(1) There are moral facts or truths, and (2) these facts or truths are independent of the evidence for them" (Brink 1989, 17). Brink notes two types of opponents of this view. The first, which includes "nihilists, emotivists, prescriptivists, and other noncognitivists" (Brink

1989, 18) consists of those who deny (1). He recognizes, however, that this is not *quite* right, as some noncognitivists believe that moral facts *do* exist, but construe them as merely psychological, social, or cultural (for examples, see Harman 1986, Blackburn 1988, and Prinz 2007). These views do not qualify as realist by the second tenet, of course. The second type of opponent consists of constructivists who affirm (1) but deny (2). Constructivism flips the scheme of justification one finds in moral realism: whereas a moral realist like Brink believes that moral beliefs are justified insofar as they accord with moral facts, a constructivist believes that moral beliefs, insofar as they are justified, constitute moral facts. This essay will focus exclusively on Brink's coherentist moral realism. His account can be broadly described as holding that moral beliefs are cognitive, truth-functional, and target moral facts which are objective and independent of our evidence for them.

### 2.2 Coherentism and Foundationalism

Even among those who believe that moral facts are independent of our evidence for them, there is disagreement about what kind of evidence justifies one's belief in a given moral fact. Suppose that you believe that it is immoral to steal candy from children. Why do you hold this belief? You might say that you believe it is immoral to steal in general, and that stealing candy from children is an instance of this more general belief. But then the question arises again: why is it immoral to steal in general? This line of questioning reveals what Brink calls the "epistemological requirement" (Brink 1989, 116). Any good account of epistemic justification for our moral beliefs must explain how our justifying beliefs (in this case, the belief that stealing is wrong in

general) are themselves justified. Brink identifies three main ways of accommodating the epistemological requirement. One can variously claim that "(1) all justification is both inferential and linear; (2) although all justification is inferential, it is not all linear; and (3) although all justification is linear, it is not inferential" (Brink 1989, 116). The two main approaches to epistemic justification in ethics are represented here: foundationalists hold (3), and coherentists hold (2).

Foundationalism holds that your moral belief is justified only insofar as it can ultimately be inferred from a basic, foundational moral belief, which is not itself justified by inference from another moral belief. It may be that your belief that stealing in general is wrong counts as such a belief. When asked why you hold this belief, you might simply say that stealing is just *obviously* wrong. In this way, foundationalism posits certain *regress stoppers*. The series of justificatory questions ends because there can be no questioning the premise that stealing in general is wrong. The belief is self-justifying and non-inferential.

Foundationalist moral realism is often called intuitionism. Moral intuitions are sometimes compared to perception of non-moral facts; as W. D. Ross says, "The moral convictions of thoughtful and well-educated people are the data of ethics just as sense-perceptions are the data of a natural science … The verdicts of the moral consciousness of the best people are the foundation on which [the moral theorist] must build" (Ross 2002, 40-41). Here there are two distinct possibilities for foundationalist theories: those that are pluralistic, and those that are monistic. Ross is an example of a pluralistic theorist. That is, he holds that there are a number of foundational moral beliefs, and that

these moral beliefs cannot be assimilated into one principle. Compare this to Robert Audi, who believes that foundational moral beliefs *can* be so assimilated with an argument from the impersonality of reason (see Audi 1990, 18). What both accounts, share, of course, is the commitment to moral facts being derivable from one or more basic foundations (intuitions in the case of Ross, practical reason in the case of Audi).

Coherentism rejects this approach to justification. Rather than attempt to find foundational beliefs on which to base all his other moral beliefs, the coherentist holds that his moral beliefs are justified insofar as they are held consistently with his other beliefs. Somewhat more formally, coherentism holds that "the justification for a moral belief *p* depends on, among other things, another moral belief *q*, whose justification consists in part in the fact that it is appropriately related to *p*" (Brink 1989, 130). The overall justificatory strategy of coherentism is one in which we "make trade-offs among the various levels of moral belief in response to conflicts among them…until we achieve maximum coherence among all our beliefs" (ibid., 131). Importantly, the degree of justification is proportional to the degree of coherence, and as such, a very incoherent moral system is far less justified than a highly coherent moral system. A perfectly coherent moral system would be considered perfectly justified.

Brink takes coherentism in ethics to be "essentially John Rawls' method of wide reflective equilibrium" (Brink 1989, 104).[1] Rawls defines wide reflective equilibrium the following way:

---

[1] I here ignore the distinction between wide and narrow reflective equilibrium, as I do not believe it is important to the case at hand. If questions arise, consider this thesis to be construing coherentism as operating through wide reflective equilibrium.

> Justification rests upon the entire conception and how it fits with and organizes our considered judgments in reflective equilibrium…justification is a matter of the mutual support of many considerations, of everything fitting together into one coherent view. (Rawls 1971, 507)

Coherence is (usually) obtained through a gradual process of reshaping one's beliefs. To illustrate this process, consider the following example. Suppose I believe that eating my beagle Herbert (under normal conditions) is wrong. In order to justify this belief using reflective equilibrium, it must cohere with my other beliefs (such as the moral belief that I should not eat my pets, the non-moral belief that Herbert is a conscious being, etc.). But suppose I hold other beliefs that are not coherent with this one. You might point out that I eat the meat of other animals, and that those animals are not different from Herbert in any morally relevant way. When faced with this contradiction, one of three things must happen if my beliefs are to remain justified. (1) I may change my opinion on meat-eating, (2) I may change my opinion on eating Herbert, or (3) I may appeal to other beliefs in an attempt to maintain coherence (perhaps I believe that dogs have souls whereas other animals do not, thus giving me a morally relevant category that Herbert satisfies and other animals do not). These are all strategies for maintaining the coherence of my system of morality, which (according to RE), justifies my holding that system. Importantly, if I simply accept the inconsistency and resort to special-pleading in Herbert's case (perhaps because I am emotionally attached to him), *all* of my beliefs lose a degree of justification. This, in short, is the method of reflective equilibrium.

### 2.3 Coherentism Does Not Imply Constructivism

As Brink is a coherentist, it is legitimate to ask whether one can be both a coherentist and a moral realist at the same time. After all, should we not see the process

of reflective equilibrium as an example of constructivism in ethics? Brink thinks we need not, and that we have good reasons for thinking of reflective equilibrium as a process of discovery, not construction.

Brink points to the fact that "the correctness of our moral beliefs appears to be independent not only of our actual justification for holding them but even of ideal justification" (Brink 1989, 31). He argues that moral beliefs (at least in an everyday context) seem to be *fact stating*. When we argue with others about moral questions, we seem to think we are getting at some kind of moral knowledge. And while knowledge implies true belief, justification does not (Brink 1989, 31). We can still ask whether someone's moral beliefs are true, even if we postulate that her moral beliefs are ideally justified. The constructivist must disallow this question, or give it a non-literal reading, because she identifies moral facts with those moral beliefs which are (perhaps ideally) justified. Constructivism, then, does not give us what we (intuitively) think morality ought to: a way for us to be justified, yet wrong.

Brink recognizes that a different version of constructivism might be able to handle this objection. We might be able to identify truth with justification "at the limit of rational inquiry" (Brink 1989, 32). That is, once we have all the non-moral facts, we can recognize that some of our previous *moral* beliefs were false. In this way, a more sophisticated version of constructivism would hold that we can recognize some moral beliefs are false, though they were justified at the time they were held. Brink is not satisfied with this, because one can still raise skeptical possibilities where justification for our moral beliefs obtains but the truth of those beliefs does not. He notes that "our

considered moral beliefs about the morality of, say, particular actions or kinds of actions play a central role in constructing and testing moral theories," and these theories in turn allow us to correct isolated errors we sometimes make in our moral judgments (Brink 1989, 34). The problem is that this method does not preclude the possibility of holding a coherent (and thus justified) system of moral beliefs, yet being systematically mistaken. Even if we imagine that we are in possession of all of the correct non-moral beliefs, this does not guarantee we will have all of the correct moral beliefs. Brink believes that our ability to conceive of such a skeptical possibility shows that we do not (even in idealized situations) think that justified moral belief entails moral knowledge.

In his consideration of Rawls's Kantian Constructivism, Brink makes another important point against constructivism in general. Rawls (according to Brink) holds that there are many competing ideals of a person (that is, what people want to be, or how they ought to be) which cannot be adjudicated, and this has implications for the truth of moral realism:

> Because of the importance of ideals of the person in moral theory and the underdetermination of ideals of the person, the truth of moral theories must be defined in terms of, or relativized to, those moral beliefs about persons upon which those theories depend (evidentially). But this implies that moral facts are evidence-dependent rather than evidence-independent. (Brink 1987, 84)

As you can see, this form of evidence-dependence derives from the underdetermination of moral theories, which in turn derives from the multiplicity of different ideals of a person. Brink notes, however, that underdetermination in the realm of theory choice only results in our inability to choose between theories, not in the evidence-dependence of the facts in question. Because the realist asserts that there are

moral facts, underdetermination only means that some moral facts may be inaccessible. Of course, if the underdetermination is severe enough, constructivism might be a better metaethical explanation than realism (Brink 1987, 84).

Even so, Brink believes that we can concede that conflicting ideals of the person underdetermine the choice of moral theories without moral theories being underdetermined *in general.* Because Brink subscribes to a coherentist moral epistemology, even ideals of the person are in principle revisable. We have no reason to suppose that ideals of the person are the only means of determining moral theory choice. If these ideals are not consistent with other considered moral beliefs, this provides us with a good impetus for changing them.

From this criticism, we might extract a more general lesson about Brink's response to constructivism. If constructivism is defended by appeal to the underdetermination of moral theory by the available facts, Brink can (in keeping with coherentism) respond by pointing to other facts that might provide this determination. Only if all the facts (moral and non-moral) underdetermine a choice of theory will we find constructivism to be a viable alternative. Even then, we might imagine Brink saying, this might not lead us to think that moral facts are evidence-dependent, but rather that we lack evidence for deciding between a few justified moral theories. Underdetermination, in short, does not imply constructivism, nor does it imply relativism.

Brink's overall line against constructivism derives from the failure of constructivism to provide good reasons for identifying moral truth with moral

justification. Not only does this position get the phenomenology of moral reasoning wrong, it fails to account for skeptical possibilities where a person is systematically mistaken. In such cases, we think that a person is *justified* but *wrong*. Moreover, even if our moral theories are underdetermined, this does not mean that the facts in question (both moral and non-moral) are evidence-dependent.

This covers the negative case against constructivism, but Brink must still provide a positive account of how evidence-independent facts operate within a coherentist epistemology. The evidence-independent, realist nature of moral facts is justified not by an alteration of the structure of coherentism, but by a suggestion about its content. He argues that "the coherence of a belief p with, among other things, realist second-order beliefs about p is evidence of p's objective truth" (Brink 1989, 131). Specifically, moral beliefs "are beliefs about moral properties and their instances (e.g., beliefs about what moral principles are true and about which actions are right). Second-order beliefs about morality, by contrast, are nonmoral beliefs about the relation between our moral beliefs and the world" (Brink 1989, 131). These realist second-order beliefs are in turn justified by their being held as part of a coherent system of beliefs.

In summary, Brink holds that coherence is not only a good method of providing justification for moral beliefs, but that coherentism is not committed to constructivism in ethics. Because Brink's account is wedded to coherentism, it is worth mentioning some of the reasons for preferring this account of justification over foundationalism. I will cover these in the next section, then move on to some of the most prominent critiques of

the account. For Brink's moral realism to succeed, there must be good arguments both against foundationalism and in defense of coherentism.

### 2.4 Three Arguments Against Foundationalism

First I will consider three arguments against foundationalism. The first is the claim that self-justifying beliefs are impossible. The second and third arguments concern moral foundationalism (intuitionism), with the second claiming that intuitionism is committed to the existence of a mysterious faculty of moral perception, and the third claiming that intuitionism is incompatible with descriptive moral relativism.

Recall that foundationalism is an instance of (3), which claims that all beliefs are justified linearly, but that not all beliefs are justified inferentially. It is this claim that allows the foundationalist to argue that there are certain self-justifying, non-inferential beliefs which form the foundation of our justified beliefs. Brink thinks that this claim is untenable, because self-justifying beliefs "can be regarded as the limiting case of circular reasoning" (Brink 1989, 16). Suppose one holds a belief *p* that is ostensibly self-justifying. Brink thinks that "in order to be justified in holding one's belief *p*, one must have reason to hold *p* to be true" (ibid.). The foundationalist needs to say that one's reason for holding *p* is *p* itself. But this obviously cannot be the case if *p* is a first-order belief, because if that were the case, *p* would target the world, not itself. Thus if *p* is a first-order belief, one must have a further belief, *q*: "My belief *p* is true." On this basis, Brink thinks that "no belief about the world can also be the reason for thinking that that belief is true" (Brink 1989, 117).

The basis for this claim might need some elaboration. Take the following ostensibly non-inferential belief: "there is a phone on the table." This is a first-order, world-directed belief. Brink's line seems to be that we cannot possibly think that such a belief is justified by merely holding it. Rather, there are a number of background beliefs at play here. For example, for this first belief to be justified, I must believe that I am a good detector of medium-sized objects in good lighting (Brink 1989, 127). If I did not have such a belief, my belief that I see a phone on the table seems to lack justification. This point generalizes: no matter what the worldly object, my belief about it seems to need justification from other beliefs about my ability to detect it. And these beliefs must always be second-order beliefs, directed at the first experiential belief.

There may be counterexamples to this general point, however. Consider this belief: "There is a phone on the table *and* I believe I am a good detector of such things in these conditions." The foundationalist might point to such beliefs as an example of a self-justifying, world-directed belief. But it does seem possible to ask for a justification of such a belief. Why should I believe both conjuncts of the proposition? Perhaps because I hold the belief, "I have seldom been deceived when I have been in conditions like this." The foundationalist could of course incorporate this belief into the first, producing another putatively self-justifying conjunctive belief. But again, we can see that even *that* belief seems to rely on further beliefs. It seems clear that such a strategy would ultimately either fall into an infinitely long conjunctive belief or some kind of recursive, non-linear conjunctive belief. It seems that Brink is correct here: any world-directed belief seems to need justification from some other belief.

Another possible alternative for the foundationalist would be to base foundational beliefs on analytic truths, which of course cannot be negated without contradiction. But there is an immediate problem here. Because analytic truths are not world-directed (the world could be completely different, yet they would still be true), it seems that the foundationalist would not get any more than obvious tautologies for foundational beliefs. If this is the case, it seems difficult to see how any synthetic beliefs could be justified, because justification from one's foundational beliefs is inferential. Therefore it seems plausible that Brink's point against the foundationalist is sound: first order beliefs must always have second-order beliefs to justify them.

The second critique of foundationalism applies specifically to theories which use foundationalism as a method of justification for *moral* beliefs. As mentioned before, foundationalism combined with realism about moral facts is often called intuitionism. Intuitionism claims that there are certain foundational moral beliefs from which we derive our other (justified) moral beliefs. As such, when we ask whether one of our moral beliefs is justified, we must be able to reduce it (possibly inferentially through other beliefs) to some foundational moral belief.

The most important historical critique of intuitionism is the claim that it posits a special faculty of the mind capable of detecting (perhaps in a manner akin to perception) the moral rightness or wrongness of a given action. Such a faculty seems quite mysterious. Despite his criticism of foundationalism, Brink is skeptical of this critique. He claims that intuitionism is not committed to the existence of any particular type of moral sense. Rather, "like nonmoral perceptual beliefs, [moral beliefs might] concern

the properties of particular people and particular events and … not result from any conscious inference" (Brink 1989, 110). In line with Brink's view about moral properties (see Brink 1989, 175), we might say that moral properties supervene on natural properties. If this is the case, the intuitionist might say that we perceive natural states of affairs and their attendant properties, and that some sets of these properties constitute certain *moral* states of affairs as well. We see the moral properties of certain natural states of affairs in the same way that we see the economic properties of certain states of affairs. Like economic theory, moral theory is just another conceptual category; it is a way of seeing the world.

The third criticism of foundationalism as a method of moral justification is more decisive. The gist of the argument is that intuitionism relies on the claim that there are self-justifying foundational moral beliefs, but that the truth of descriptive moral relativism entails that these beliefs do not exist. First we must get clear on what descriptive moral relativism entails. This is the claim that there is radical disagreement about morality. This is a fairly uncontroversial thesis, even for moral realists; it is difficult to deny that different cultures have different moral rules, and that what counts as permissible in one culture is often impermissible in another. Second, we should be clear that this argument only applies to foundationalists who claim that foundational moral beliefs lead to moral knowledge which is objective and cross-cultural. Foundationalists who accept some form of relativism are not affected by this argument.

With these two points in mind, let us consider how the foundationalist would respond this argument. Brink notes that there is one common response: one can "appeal

to the principle of noncontradiction to deny the existence of conflicting moral intuitions … People may hold conflicting moral views quite firmly, but these moral views need not be foundational" (Brink 1989, 111). The intuitionist might even claim that certain self-justifying beliefs are not immediately self-evident. Rather, they are mediately self-evident: we may need to subject our beliefs to a thorough dialectical investigation to determine which beliefs are in fact self-evident (Brink 1989, 112).

There are a number of problems with this claim. First, one may doubt that the outcome of such a dialectic procedure will arrive at self-evident moral beliefs. Without going into a detailed anthropological survey, it still seems fair to say that intercultural (and even interpersonal) moral disagreement runs quite deep. There is a real danger of the intuitionist simply reaffirming the beliefs which are strongly embedded in her culture and upbringing. The intuitionist's position thus depends on the questionable claim that a dialectical procedure would eventually result in self-evident moral beliefs. If it does not, she must claim that the procedure has gone awry somewhere. Second, one may ask how we know that these supposedly self-evident moral beliefs are in fact true; simply claiming they are self-justifying is not necessarily enough to preclude this question. Here, the intuitionist must simply deny that such a question is possible. But it is difficult to determine a reasoned basis for such a denial, especially if (as seems likely) there remains strong disagreement at the end of the dialectical procedure, or if there is disagreement about how the dialectic is to proceed in the first place.

Intuitionism, then, seems like a difficult position to hold. Because every world-directed belief must be a first-order belief, it is not obvious that self-justifying moral

beliefs can exist at all. Moreover, it is even less obvious that we would be able to determine exactly what these beliefs might be. If a coherentist account of the justificiation of moral beliefs is possible, it will most certainly avoid the first objection (and possibly the third as well). But there are a few important arguments against coherentism that we must also consider before endorsing such a position.

## 2.5 Three Criticisms of Coherentism

There are three important criticisms of coherentism: the regress argument, the isolation objection, and the multiple coherent systems (MCS) objection. I will consider each of these in turn, though two (moral) species of the latter objection will occupy much of the third section. I contend that the regress argument and the isolation objection fail, but that the MCS objection is more difficult to deal with.

We will consider the regress argument first. Recall the three options mentioned earlier in accommodating the epistemological requirement: "(1) all justification is both inferential and linear; (2) although all justification is inferential, it is not all linear; and (3) although all justification is linear, it is not inferential" (Brink 1989, 116).[2] The regress argument claims that (1) involves a vicious regress, and that only foundationalism can stop such a regress. Moreover, it claims that (2) (which represents coherentism) "succeeds only in trading the vicious regress in (1) for vicious circularity" (Brink 1989, 105).

To see why (1) ("all justification is both linear and inferential") might lead to a vicious regress, we need only consider its two requirements. First, the requirement of

[2] "Linear" justification is justification which never loops back on itself.

linearity holds that if a belief P is justified by another belief Q, that belief Q cannot (even through a series of other beliefs) be justified by belief P. The linearity requirement thus excludes all possible cases of inferential loops (and such loops are necessary for coherentism). Second, because all justification must be inferential, this excludes non-inferential, self-justifying beliefs on which foundationalism relies. So, because (1) rules out circular inferences and non-inferential justification, it does not provide a way of ultimately justifying beliefs. Under (1), when we ask why our belief P is justified, we may say that it is justified by another belief Q. But because Q cannot be self-justifying, and cannot be justified by P, it must be justified by a different belief, R, which in turn must be justified, and so on. If we accept (1), it seems we are required to provide an infinite number of beliefs in order to obtain justification. This is surely implausible, not only because it seems psychologically impossible that we could have an infinite number of beliefs, but because even if we did, it does not seem we could ever actually get justification from such an infinite series. This regress is vicious because it can never obtain the justification it seeks to provide.

Of course, coherentism rejects (1) as well, so the key part of the argument in favor of foundationalism is the claim that (2) represents a case of vicious circular reasoning. Brink believes that this claim can be overcome. He notes that there are two different kinds of epistemic justification: contextualist and systematic. Systematic justification "results from consistently applying the epistemological requirement that justifying beliefs be justified" (Brink 1989, 123). Contextualist justification is "partial or incomplete justification and results from refusing to apply the epistemological

requirement consistently" (Brink 1989, 123). Brink believes that we justify our beliefs in everyday life using contextualist justification. We do not thoroughly apply the epistemological requirement (that every justifying belief itself be justified), and as a result, the process of justification we engage in is generally short and linear. When I am asked why I believe stealing candy from children is wrong, it is not usually expedient for me to do more than justify my belief based on the belief that stealing in general is wrong. I do not (unless I am philosophizing) ask why stealing in general is wrong. This psychological bias leads to people to believe that systematic justification will look something like contextualist justification (that is, short and linear!). Brink thinks this is not the case. Rather, he contends that if I were to question the justification of my justifying beliefs (and thus ask for systematic, rather than contextualist justification), I would have to bring in many background beliefs that are themselves only justified because they are held as a coherent system. This is the strategy at work in (2), where the linearity requirement for justification is dropped. As Brink argues, "there is no clear reason to suppose that linearity can or need be preserved when the epistemological requirement is consistently observed and the demand for justification (as a result) becomes systematic and not merely contextualist" (124).

Here it is not clear that Brink has thoroughly avoided the regress argument, though he has given us a deflationary explanation for why we might initially lean toward an account of justification that preserves linearity. Brink's argument here involves pointing out that we have a bias in favor of linear accounts of justification, but the presence of such a bias does not *prove* that justification can be non-linear. What the

coherentist needs is an account which shows that non-linear (circular) justification need not be vicious.

Keith Lehrer provides one way of doing this. He claims that one should only consider a belief justified if one also believes that the first belief "is a trustworthy guide to truth" (Lehrer 2000, 137-138). Of course, we might see a regress here: I believe that the first belief is a trustworthy guide to truth, but why do I believe this second belief is the same? This amounts to applying the epistemological requirement. Lehrer avoids the regress by claiming that we can fall back on the following belief: "I am trustworthy (worthy of my own trust) in what I accept with the objective of accepting something just in case it is true" (ibid.).[3] Of course, we can (and if we are after systematic justification, we must) ask why this belief is itself justified. But Lehrer believes that it is justified because of its reflexive nature. If I believe that I am trustworthy, then that trustworthiness reinforces the belief that I am trustworthy (Lehrer 2000, 142). This position is obviously circular, and Lehrer admits as much. But he claims that it is not *viciously* circular:

> If we have a principle that explains why it is reasonable to accept what we do, it is a virtue rather than a vice that it should at the same time explain why it is reasonable to accept the principle itself. The other alternative is that the principle should be a kind of unexplained explainer. (Lehrer 2000, 143)

Appealing to this kind of reflexive belief gives us a principled way to avoid the charge of circular reasoning, and I contend that this response (though quirky) stops the regress argument in its tracks. In fact, we will see that the regress argument is certainly not the strongest argument against coherentism.

---

[3] Here I will not get into the technical difference between "accepting" a belief and considering it justified.

The second criticism of coherentism is commonly referred to as the isolation objection. It is often put metaphorically: coherentism cannot provide the "contact with reality" necessary to be a good theory of justification. Coherentism holds that beliefs are justified insofar as they are held as part of a coherent system, but this says nothing about whether such beliefs are held as a result of experience. The fact that coherentism requires all beliefs to be inferentially justified seems to be a problem, because beliefs based on experience ("experiential beliefs") *do not* seem to be held inferentially.

Laurence BonJour provides us with one possible response: he distinguishes two different senses of the word "inferentially." First, we may use this word to describe how we arrive at a given belief. Second, we may also use this word to explain how a belief is justified. This distinction is important, because evidence from our senses is generally considered to be non-inferential. BonJour then goes on to explain that the coherentist can easily admit that certain experiential beliefs are *arrived at* non-inferentially, but *justified* inferentially. He argues that the *origin* of a belief (how we arrive at it) is simply not relevant to whether this belief is justified (BonJour 1985, 113).

BonJour's point seems intuitively plausible, for denying it would seem to constitute an instance of the genetic fallacy. We should be careful in saying it is *irrelevant,* however. Take the following example: suppose that Nazi scientists were the first to see (in experiment) that rockets with a particular type of fuel traveled the longest distance. It is obviously fallacious to assert that we have no reason to believe what these scientists discovered *because they were Nazis*. But it is of course right to think that the justification of this belief might be affected by the conditions which surround its origin.

Perhaps the Nazi scientists in question were under a great deal of political pressure to produce a long distance rocket, and so inflated their measured flight distances. If we have reason to believe this, we have reason to doubt their discovery.

Analogously, the coherentist should admit that the circumstances surrounding the origin of a belief may affect the plausibility of that belief. But this does not commit him to saying that experiential beliefs, arrived at non-inferentially, must also be justified non-inferentially. Rather, it gives him a good reason to think the opposite. After all, it is typically the *foundationalist* who thinks that experience gives us self-justifying, non-inferential belief. Rather, the coherentist can hold that, "observational belief always depends on the general knowledge that beliefs of [a] specific kind are nomologically reliable indicators of the actual presence of the sort of factual situation whose existence they assert" (BonJour 1985, 116). This general knowledge might be cashed out as a conjunctive proposition like "I accurately perceive things in quiet, well-lit rooms *and* I am not colorblind *and* I accurately perceive things when I am not sleepy…"

Of course, such a general body of knowledge is only considered knowledge by a person because it is (on the standard account of knowledge) true justified belief.[4] And this general belief is itself justified by its coherent fit with the rest of the system. In this way, the coherentist can give an account of how non-inferential, observational belief plays a part in one's system of belief.

The third important argument against the coherentist is perhaps the most challenging: the MCS objection. This objection can be raised in a number of ways; later

[4] I am of course leaving out Gettier cases, but I do not think these affect the case at hand.

I will consider two specifically *moral* MCS objections, but here I will raise the objection as an expansion of the isolation objection. Remember that the isolation objection claims that coherentism cannot provide the intuitively necessary connection between justification and experience. Because all beliefs must be justified by inference, it appeared that non-inferential experiential beliefs could not be justified. Now even if we adopt BonJour's solution to this problem, we are still faced with the possibility that someone might reject all experiential beliefs but still have a coherent system of belief (call such a system an "isolated system"). This would obviously be starkly different from a system of belief which did allow some experiential beliefs. And this is where the MCS objection arises: both of these systems would be equally justified. The coherentist must thus claim, not only that we *can* adopt a system of beliefs which incorporates experiential beliefs, but that we *must* adopt such a system.

BonJour can rule out *some* intuitively unjustified systems of belief by further specifying the requirements of coherence. To this end, he proposes five conditions:

> (1) A system of beliefs is coherent only if it is logically consistent.
> (2) A system of beliefs is coherent in proportion to its degree of probabilistic consistency.
> (3) The coherence of a system of beliefs is increased by the presence of inferential connections between its component beliefs and increased in proportion to the number and strength of such connections.
> (4) The coherence of a system of beliefs is diminished to the extent to which it is divided into subsystems of beliefs which are relatively unconnected to each other by inferential connections.
> (5) The coherence of a system of beliefs is decreased in proportion to the presence of unexplained anomalies in the believed content of the system. (BonJour 1985, 95-99)

The concept of "anomaly" at work in (5) is "a fact or event…which is claimed to obtain by one or more of the beliefs in a system…but which is incapable of being

explained…by appeal to the other beliefs in the system" (BonJour 1985, 99). BonJour appeals to a common Hempelian account of scientific explanation, in which "particular facts are explained by appeal to other facts and general laws from which a statement of the explanandum fact may be deductively or probabilistically inferred" (ibid., 98). On this view, a system of beliefs which includes a number of anomalies (which cannot be explained in a Hempelian fashion) is less coherent than a system which includes fewer anomalies. We can thus see that BonJour's account is far more than a mere matter of holding a system of beliefs which is coherent (where coherence is itself an unanalyzed concept). It is rather a detailed view which incorporates a specific concept of explanation.

Unfortunately, while these conditions spell out the specific conditions of coherence, they do not on their own rule out an isolated system. This is because one could imagine an isolated system where many different beliefs played an inferential role in other beliefs, where there were no relatively unconnected subsystems, and there were no unexplained anomalies, but where experiential beliefs are still banished. Such a system would certainly have to be somewhat fanciful, but if it is just as justified as a serious, commonsense system, this is still a problem.

BonJour believes that we can respond to this type of objection, but only by introducing what he calls the observation requirement (OR): "a cognitive system … must contain laws attributing a high degree of reliability to a reasonable variety of cognitively spontaneous [experiential] beliefs" (BonJour 1985, 141). This requirement would independently rule out an isolated system.

I do not think that adding this requirement will do the trick. This is because if we want systematic justification, there is nothing preventing us from asking what justifies the OR in the first place. If the OR is only justified insofar as it is held as a coherent system (perhaps in accordance with BonJour's five conditions), it seems difficult to see why someone who holds an isolated system would accept it. Here BonJour has two options: he could either accept that some justified isolated systems do not include the OR, or he could assert that the OR has some special status among other beliefs. Neither option looks acceptable to a coherentist. What we need here is a way to incorporate the OR into *any* justified coherent system without making the OR itself a foundational belief.

I think this problem can be overcome, and that we need not look to foundationalism for an answer as to how we might preclude an isolated system. In discussing the observation requirement, BonJour hints at one possible avenue for this kind of account: one must not only have a coherent system of belief, but be able to explain *why* that system is coherent (BonJour 1985, 98-100). As mentioned before, BonJour relies on a standard Hempelian account of explanation. I propose that with a better account of explanation, we can solve this version of the MCS objection without modifying BonJour's conditions or accepting foundationalism.

### 2.6 Kitcher: Explanation as Unification

In his article "Explanatory Unification," Philip Kitcher provides an account of scientific explanation which is both thoroughly compatible with coherentism and which will allow us to avoid the observation requirement. First he points to what he calls the

"official view," which is the standard Hempelian account of explanation. On this view, a theory is explanatorily relevant when "the explanatory information adduced affords good grounds for believing that the phenomenon to be explained did, or does, indeed occur" (Hempel 1991, 318). What constitutes good grounds for believing that a phenomenon occurred is a deductive-nomological or inductive-statistical argument, where the former invokes general scientific laws and auxiliary facts to logically deduce the predicted result, and the latter invokes statistical laws (along with auxiliary facts) to provide a high probability that the predicted result will occur. An event is explained, according to Hempel, when we can provide a good argument for its occurrence.

Kitcher believes that lurking behind this view is another, unofficial view, which Hempel himself expressed: "What scientific explanation, especially theoretical explanation, aims at is not an intuitive and highly subjective kind of understanding, but an objective kind of insight that is achieved by a systematic unification, by exhibiting the phenomena as manifestations of common, underlying structures and processes that conform to specific, testable, basic principles" (Kitcher 1991, 330). Kitcher believes that we can understand explanation as unification itself.

To get at how this might be so, Kitcher explains how we draw up explanatory arguments in the first place. In his account, we have a set of arguments available for explanatory purposes (what he calls an "explanatory store") (Kitcher 1991, 332). The crucial feature of his account is that he is attempting to specify "conditions which must be met by the explanatory store" so as to rule out explanatory stores which are merely "unrelated individual arguments which can be used in individual acts of explanation"

(ibid.). More formally, he calls our explanatory store E, and the set of accepted sentences (describing phenomena) K. The task for the unificationist is to explain the relation E(K) ("the explanatory store over K") which is "the set of arguments acceptable as the basis for acts of explanation by those whose beliefs are exactly the members of K" (Kitcher 1991, 332).

Kitcher believes that we can define good conditions for E(K) by attending to the argument patterns employed in E. When a theory "provides one (or more generally, a few) pattern(s) of argument which can be used in the derivation of a large number of sentences which we accept," that theory is unified (and thus explanatory). As an example of this kind of argument pattern, Kitcher points to Darwin, who explained the features of organisms by natural selection. Darwin's account is highly explanatory because the argument pattern for all species is the same: for any organism O and any given characteristic *x* (anatomical, behavioral, etc.) of O, O will have *x* iff *x* improves the fitness F of O. To flesh out this example, suppose the set K of sentences describing observed phenomena includes (among other things) two sentences: "Giraffes have long necks" and "Male frogs croak loudly to attract mates." Darwin's account is explanatory because both of these sentences are explained with one argument pattern: the giraffe has a long neck because those giraffes with longer necks had greater fitness, and male frogs croak loudly because the louder the frog was, the more likely it was to attract a mate, also increasing its fitness. This explanation is obviously better than (for instance) an explanation in which the loudness of male frogs' croaks is explained by their need to exercise their throats and the giraffes' long necks are explained by their ancestors' neck-

stretching. Of course, Darwin's explanation also commits him to applying the argument pattern the same way in all relevantly similar cases. So, given that (say) complex organs are a characteristic of many organisms, Darwin must explain how complex organs improve the fitness of an organism, and be able to show how such organs could have developed via incremental changes (across *many* generations) from simple to more complex organs (Kitcher 1991, 334).

To give us a better grasp of what an argument pattern is, Kitcher further formalizes the concept. An argument pattern "is a triple consisting of a schematic argument, a set of sets of filling instructions containing one set of filling instructions for each term of the schematic argument, and a classification for the schematic argument" (Kitcher 1991, 335). First, a schematic argument is one where the non-logical terms of an argument have been replaced with dummy letters. Second, the filling instructions are directions on how to replace each dummy letter with a non-logical term which occurs in a particular sentence of our set of accepted sentences K. Finally, a classification is "a set of sentences which describe the inferential characteristics of the schematic argument" (ibid.). The classification tells us "which terms in the sequence are to be regarded as premises, which are to be inferred from which, what rules of inference are to be used, and so forth" (ibid.).

Furthermore, there are a number of conditions which Kitcher believes must obtain for a sequence of sentences to constitute an argument pattern:

(i) The sequence has the same number of terms as the schematic argument of the general argument pattern.

(ii) Each sentence in the sequence is obtained from the corresponding schematic sentence in accordance with the appropriate set of filling instructions.
(iii) It is possible to construct a chain of reasoning which assigns to each sentence the status accorded to the corresponding schematic sentence by the classification. (Kitcher 1991, 335).

Returning to the previous example, we can see more clearly how Darwin's argument pattern would be formalized by Kitcher. Any sequence of sentences which we wish to call an evolutionary explanation must meet conditions (i)-(iii). Moreover, it must be modeled on a schematic argument which I give below. Note that the argument pattern adduced below is given meaning by the filling instructions; without them, it is trivial:

(1) $F = N_a/N_b$.

(2) $Ot \leftrightarrow F > 1$.

(3) $\sim Ot \leftrightarrow F < 1$.

(4) $F > 1$.

(5) $Ot$.

This schematic is the general form of the argument, but what truly does the work here are the filling instructions.[5] They tell us that O is to be replaced by a species of organism, *t* is to be replaced by a characteristic of that organism, F is to be replaced by the organism's absolute fitness. They further tell us that $N_a$ is to be replaced by the number of organisms with a given genotype *after* a selection, and $N_b$ is to be replaced by the number of organisms of a given genotype *before* a selection. The classification tells us that O ranges over *t* with the one-place relation "has" (organism O has characteristic

[5] See Kitcher (1991) for another example of schematic arguments with filling instructions.

*x*), and that (1)-(4) are to be used as premises while (5) is to be used as a conclusion. Of course, this example schematic argument is *far* too simplistic to actually model an evolutionary argument, partially because absolute fitness is not always easy to measure (meaning it may not cleanly match up with our set of accepted sentences K). Moreover, it does not cleanly distinguish between those characteristics of an organism which lend toward its increased fitness and those traits which are merely "along for the ride" but do not increase its fitness. My aim here is not to provide a precisely accurate Darwinian schematic argument, but to give a good example of a schematic argument in the first place.

While having a small number of argument patterns does give a theory more unified power, this is not the only requirement. The argument patterns must also be stringent. The stringency requirement is formed by "(1) the conditions on the substitution of expressions for dummy letters, jointly imposed by the presence of non-logical expressions in the pattern and by the filling instructions; and, (2) the conditions on the logical structure, imposed by the classification" (Kitcher 1991, 336). Thus an argument pattern which failed to provide detailed conditions on either count would not be considered stringent. This means that a theory which employs very few argument patterns but has either (A) very lax conditions for the substitution of dummy letters or (B) very lax conditions on the logical structure would fail to be unified. Taken together, the number and stringency of argument patterns are considered the two conditions which measure a theory's explanatory power.

Returning once again to our example, we can see that these two conditions are fairly good at capturing our intuitive notion of explanatory power. Consider the previously mentioned alternative to the evolutionary explanation, which has different explanations for why male frogs croak loudly and giraffes have long necks. The kind of argument pattern instantiated by the first explanation might be the following:

(1) $Ot \rightarrow Ol$.

(2) $\sim Ot \rightarrow \sim Ol$.

(3) $Ot$.

(4) $Ol$.

In this case, the filling instructions would specify that *t* is the characteristic of needing to exercise one's throat, *l* is the characteristic of croaking loudly, and O is any organism which is a frog. The classification tells us that O is a one-place relation ("has") ranging over *t* and *l* and that (1)-(3) are to be used as premises and (4) is to be used as a conclusion following from these premises.

It is easy to see that the argument pattern instantiated by the second explanation about the long necks of giraffes will be somewhat different. Though the logical structure will remain the same, the filling instructions and classification will be significantly different: *t* and *l* will be "needing to stretch one's neck" and "having a long neck", and O will refer to any organism that is a giraffe (instead of a frog). Kitcher's point should be quite obvious here: if we are faced with one explanation for a number of phenomena or many explanations for many phenomena, we should prefer the one explanation.

## 2.7 Unificationist Epistemology

I contend that we can use Kitcher's unificationist account of explanation to generate a stricter approach to coherentism which is not vulnerable to the MCS objection adduced above. Before I proceed, a few quick notes are in order. First, recall that an isolated system is one where experiential beliefs are entirely excluded. Here it is important to make the point that even someone who holds an isolated system will have observational input. What makes such a system isolated is not that the person does not hear, see, or feel anything, but that she does not think that any of these experiences justify her in believing the content of that her senses deliver to her. She rules out experiential beliefs not on the basis of a lack of observational input, but on the basis of skepticism about her perceptions.

Second, as I will be deploying Kitcher's work for epistemological ends, it is necessary that we see the set of accepted sentences K as containing sentences not only about things in the world, but also sentences about one's beliefs about things in the world, and even sentences about one's beliefs about one's beliefs about things in the world. On this reading, K would include sentences like "humans have ten toes" *and* sentences like "I believe humans have ten toes" and "I believe that I believe that humans have ten toes." I do not think this reading poses a problem for Kitcher's account, though it does depart significantly from its original aims. From here, the strategy will be to ask how the hypothetical proponent of an isolated system would explain her entire set of accepted sentences K. If an isolated system cannot explain its set of accepted sentences

(including the statements describing its beliefs) as well as a non-isolated system, this provides a good means of calling the isolated system less justified.

With these points in mind, I will consider two different types of isolated systems in turn: many-pattern and single-pattern isolated systems. Our account of coherentism must be able to deal with both kinds of isolated systems if it is to avoid the MCS objection. A many-pattern isolated system would involve many different argument patterns which serve to explain why different experiential beliefs are rejected. For instance, someone could hold that she does not accept what she sees because she believes that there are small gnomes manipulating her optic nerve, and does not accept what she tastes because her saliva has been chemically modified, etc. In this case, her higher-order belief that experiential beliefs should be rejected is really based on a number of different argument patterns with (presumably) different filling instructions and classifications. It is easy to see that when compared with the standard belief about the role of experiential beliefs, this kind of isolated system would not be as unified. That is, someone who holds the (presumably reasonable) belief that he is a good detector of medium-sized dry goods in good lighting can explain his experiential beliefs using just one general argument pattern:

(1) $\forall y(By) \rightarrow \forall x(Cx)$.

(2) $\forall y(\sim By) \rightarrow x(\sim Cx)$.

(3) $\forall y(By)$.

(4) $\forall x(Cx)$.

In this argument pattern, the filling instructions would tell us that *y* is to be replaced by any statement about an experiential belief and *x* is to be replaced by any statement about a second-order belief whose content endorses *y*. Moreover, the instructions would tell us that B is a one-place relation ("is formed under good epistemic conditions": decent lighting, sobriety, etc.) ranging over *y*, and C is a one-place relation ("is epistemically warranted") ranging over *x*. The classification, once again, tells us that (1)-(3) are to be used as premises and (4) as a conclusion. Not only does this system rely on just one argument pattern to explain why it accepts experiential beliefs, this argument pattern is stringent. Its filling conditions do not let just any statement in; more specifically, they exclude statements about beliefs which are not formed under good epistemic conditions.

Now, a system of beliefs which instantiates this argument pattern will obviously be more unified than the aforementioned many-pattern isolated system which instantiates many different argument patterns with many different filling instructions and classifications. As such, this case is relatively easy. However, this is not the end of the story, as there is a second kind of isolated system which we must dispense with: the single-pattern isolated system.

This kind of isolated system would reject experiential beliefs using a single argument pattern. We might represent this possibility by imagining someone who believes in the existence of a Cartesian evil demon: she has the higher-order belief that her experiential beliefs have been thoroughly warped by the demon's influence. As a result, she rejects all of her experiential beliefs.

Kitcher considers a similar problem he calls "spurious unification." We might imagine this kind of isolated system as a kind of argument pattern which unifies all of one's beliefs by "completely deriving all of them using arguments which instantiate one pattern" (Kitcher 1991, 342). This seems like a reasonable construal of the isolated system in question because every belief such a person would have is explained by the evil demon's tampering. This kind of argument pattern might look like the following:

(1) $\forall x(Dx)$.

(2) $\exists y(\sim Dy)$.

The filling instructions for this pattern would specify that x refers to all but one accepted sentence in K, D is a one-place relation ("is an illusion caused by the evil demon") ranging over x, and y is one accepted sentence which states: "the evil demon exists." (We should note that because we are here construing K as including statements about beliefs, without (2) we would see the paradoxical consequence that the belief "the evil demon exists" is itself an illusion caused by the evil demon.) This kind of isolated system appears to explain all of its accepted sentences with a single argument pattern.

Kitcher has a ready reply about such a universal argument patterns: "they fail dismally when judged by the criterion of stringency" (Kitcher 1991, 342). Stringency, again, is a combination of two factors: "(1) the conditions on the substitution of expressions for dummy letters…[and] (2) the conditions on the logical structure" (Kitcher 336). It seems clear that the argument pattern above fails on the first count; it is "very lax in allowing any vocabulary whatever to appear in the place of" *x* (Kitcher 1991, 342). As in our Darwinian example, the logical form of an argument schematic is

often trivial or uninteresting; the filling instructions do most of the work in defining the argument pattern. For this reason, the filling instructions must be quite specific. Non-stringent argument patterns allow far too broad a range of objects to occupy places in the non-logical vocabulary. Because such an argument pattern is not stringent enough, it is not unified, and because it is not unified, it is not explanatory.

There is another important note to be made about this kind of argument pattern. Though it purports to explain all of one's beliefs by calling them illusions created by the evil demon, it is unclear exactly how the belief in the evil demon itself could be explained. This becomes especially problematic when one considers the fact that the argument pattern above rejects *all* beliefs but the one which states that the evil demon exists. As a result, there is no clear argument pattern left which one might point to in explaining one's belief in the evil demon.

Consider the non-isolated system's favored argument pattern (mentioned above). This argument pattern explains one's experiential beliefs with reference to the epistemic conditions under which those beliefs are formed. It is easy to see that the non-isolated system would have to explain another (second-order) belief (B) about the connection between the set of epistemic conditions asserted about *x* and the epistemic warrant assigned to them in *y*. (Construe B as the following: "an experiential belief formed under good epistemic conditions is a warranted belief"). But this is not a problem for the non-isolated system, because this belief may be explained by any number of other beliefs that the non-isolated system has at its disposal. Avoiding a regress, we might even say that this iterative process of explanation could eventually bottom out in

something like Lehrer's reflexive belief: "I am trustworthy (worthy of my own trust) in what I accept with the objective of accepting something just in case it is true" (Lehrer 2000, 137-138). And here we need only assert that just as this belief can provide warrant for itself, it can ultimately explain itself as well. This route of explanation is not available to the single-pattern isolated system because such a system discounts all beliefs as illusions created by the evil demon, and so, again, has no resources with which to explain the belief in the evil demon.

There is another potential route for the proponent of an isolated system, and this represents a middle-way between the many-pattern and single-pattern system. She may do something analogous to "artificially introduc[ing] restrictions on the pattern to make it more stringent" (Kitcher 1991, 342). In our case, this means she might turn to the very same argument pattern which the non-isolated system employs, but change the filling instructions. This pattern would (again) look like this:

(1) $\forall y(By) \rightarrow \forall x(Cx)$.

(2) $\forall y(\sim By) \rightarrow x(\sim Cx)$.

(3) $\forall y(By)$.

(4) $\forall x(Cx)$.

The filling instructions she would introduce might specify that *y* is to be replaced by any statement about an experiential belief and *x* is to be replaced by any statement about a second-order belief whose content rejects *y*. The instructions would further tell us that B is a one-place relation ("is formed under 'good' epistemic conditions") ranging over *y*, and C is a one-place relation ("is epistemically *un*warranted") ranging over *x*. This kind

of argument pattern appears to be stringent, and could be used to explain just as many statements about experiential beliefs as the similar argument pattern employed by someone who holds a non-isolated system.

This strategy runs into a problem similar to the one encountered earlier: what explains one's second-order belief (B`) about the connection between the set of epistemic conditions asserted about *x* and the lack of epistemic warrant asserted about them in *y*? (The content of B` would be: "an experiential belief formed under 'good' epistemic conditions is an unwarranted belief"). If such a belief is explained with reference to some other higher-order belief which denies the epistemic warrant of beliefs *generally*, then an isolated system which uses this argument pattern is in just as bad a shape as the single-pattern system mentioned earlier. Alternatively, this system might avoid such a general belief and propose a higher-order belief (B``) that *only* targets B`. But then we run into a regress, for what justifies B``? An isolated system cannot explain itself with reference to a version of Lehrer's reflexive belief, because a belief which stated: "I am *not* trustworthy (worthy of my own trust) in what I accept with the objective of accepting something just in case it is true" would not be self-explaining in the same way.

It appears that there is no good argument pattern which makes an isolated system as good at explaining its own coherence as a non-isolated system. Because at least some non-isolated systems *do* explain why they are coherent, I think we can build a requirement which rules out isolated systems without relying on the observation requirement. This can come in the form of a sixth condition on the coherence of a

system (in addition to the five that BonJour mentions): (6) The coherence of a system of beliefs is increased to the extent that it can provide a explanation for *why* it is coherent, with more unified systems being more explanatory and less unified systems being less explanatory. (Construe "unified," a la Kitcher, as some ideal combination of stringency and few argument patterns).

With this condition, we can see that there is at least a plausible case to be made that the MCS objection can be overcome without the use of the (problematic) observation requirement. We should not be overly hasty: there may be ways of conceiving of an isolated system which are not preempted by this sixth condition. But I can think of none which are not either a many-pattern or single-pattern isolated system (or something in between). If there are indeed none, we will have a good reason for thinking that the MCS objection fails.

There is one last objection that we must anticipate, and it relates to the whole strategy that both BonJour and I are attempting: if I am designing an account of coherentism which excludes all cases of isolated systems, is this not merely an ad hoc addition to the concept of coherence? To this I can only remind the reader of the dialectic at work here. Remember that the MCS objection operates by pointing to an intuitively unjustified system and claiming that coherentism licenses it. Because such a system could not possibly be justified, the argument claims that coherentism is false. The argument is thus a reductio ad absurdum. But if we suitably adjust our account of coherence so that it precludes these kinds of systems, the reductio does not go through. It would certainly be a problem to simply stipulate that isolated systems are not justified,

as such a move would simply beg the question in favor of coherentism. But that is not what we have done here. Rather, I have provided an account of coherentism which both excludes isolated systems and uses an account of explanation which is independently plausible (see Kitcher (1991) for the many independent virtues of this account of explanation).

### 2.8 Section Conclusion

Let us return to the implications of these arguments for Brink's moral realism. As we have seen, there is good reason to believe that foundationalism has serious problems. It is not clear that self-justifying beliefs exist. Furthermore, when combined with moral realism, foundationalism runs the risk of simply justifying the preexisting beliefs of one's upbringing and culture. It is not clear that even a careful investigation will succeed in arriving at self-evident moral beliefs, and it is furthermore not obvious that such beliefs need no justification. Out of the three criticisms of coherentism, the first two are fairly easily dealt with. Though coherentism requires circular justification, such justification is not *viciously* circular. Additionally, coherentism is compatible with input from the external world because it can differentiate the causal source of a belief from the justification of that same belief. The MCS objection is a much more difficult problem, and one that I believe deserves a great deal more space. I believe that we can avoid the non-moral MCS objection above by including an explanatory requirement in coherentism. However, I do not think that all species of the MCS objection are so easily dealt with. In the coming section I will propose two additional MCS objections that specifically arise for *moral* systems of belief.

# 3. INTRANSIGENCE: TWO NEW MCS OBJECTIONS

In this section I will draw from the work of Jonathan Haidt and Richard Lazarus in sketching a cognitivist critique of coherentist moral realism. In the first subsection, I argue against constitutive cognitivism and defend a form of etiological cognitivism that claims that intuitive moral judgments cause moral emotions. In the second subsection, I will review the evidence for the social intuitionist model of moral decision-making and argue that when combined with the CAD triad hypothesis, it implies that reasoned moral judgments are caused by intuitive moral judgments. In subsection three, I will argue that this view presents two problems for coherentism: "the problem of intransigence" and the "limited problem of intransigence." These problems represent two new versions of the MCS objection. Finally, in subsection four I analyze what these objections mean for Brink's coherentist moral realism and conclude that there is as yet no adequate response to such objections.

## 3.1 Constitutive or Etiological Cognitivism?

Here I will consider two possibilities for the relationship between judgments and emotions:

*Constitutive Cognitivism*: judgments constitute emotional states.

*Etiological Cognitivism:* judgments cause emotional states.[6]

This thesis takes up etiological cognitivism for two main reasons. First, we have good and independent reasons for believing it to be the case. Second, it is important to the critique I develop later that we have good reasons to believe etiological cognitivism to

[6] See Scarantino (2010) for more about these distinctions.

be true. If it were not, we would risk getting mired in the seemingly endless debate over cognitivism and non-cognitivism. I believe this debate can be circumvented with the right understanding of the relationship between judgments and emotions.

The main target of my critique of constitutive cognitivism is Martha Nussbaum, who holds that "emotions always involve thought of an object combined with thought of the object's salience or importance; in that sense, they always involve appraisal or evaluation" (Nussbaum 2001, 23).[7] Here it is important to note that Nussbaum has a fairly liberal conception of what judgment is. Judgment is not necessarily "the presence of elaborate calculation, of computation, or even of reflexive self-awareness" (ibid.).

Nussbaum thinks that there are good reasons for identifying emotions with a species of judgment. For one, most emotions, like judgments, appear to have intentionality. They are *about* their object. To elaborate on this, we should turn to the work of Anthony Kenny, who distinguishes two senses of the word "object": material and formal. As he puts it, "Anything which can be [φed] is a material object of φing," while "[t]he formal object φing is the object under that description which must apply to it if it is to be possible to φ it" (Kenny 2003, 132). Nussbaum holds that the object of an emotion is what that emotion allows one to perceive. Using Kenny's language of formal objects, emotions (according to Nussbaum) perceive their material objects. Thus when speaking about the death of her mother, Nussbaum says that, "my fear *perceived* my mother both as tremendously important and as threatened" (emphasis added) (Nussbaum

[7] Throughout this paper I will use the words "cognitive appraisal" and "judgment" interchangeably. The former appears to be preferred by psychologists in order to be very specific about what they mean; "judgment" may contain too much philosophical baggage for empirical work. Though I use the terms interchangeably, I intend for the meaning of "judgment" to simply be "cognitive appraisal" and nothing else.

2001, 27). Her mother's precarious condition is the material object of Nussbaum's fear simply because it can be feared. And her mother's condition constitutes a material instantiation of the formal object of fear itself, which is simply "danger" (or perhaps "a danger of loss or harm").

In this framework, it is also easy to see that different emotions can embody different kinds of value judgments about their objects. Even when the object stays the same, one's emotions toward an object might change over time as one's perceptions about the object change. For instance, someone who fears spiders might undergo systematic desensitization therapy to get rid of this fear. If the therapy is successful, she would see the same object (a spider) with different emotions after the therapy (perhaps even admiration!). What is important here for Nussbaum's view is that even though our emotions about an object may change, those emotions will always constitute some *judgment* about that object.

Moreover, Nussbaum thinks that having certain beliefs is both a necessary and a sufficient condition for having certain emotions (Nussbaum 2001, 34). Nussbaum is thus committed to the position that one's value judgment about an object cannot change without an accompanying change in one's emotions toward the object, and vice versa.

The claim that certain judgments are necessary for certain emotions seems difficult to defend. This is because it opens her position up to a fairly obvious counterexample: suppose a man, John, grew up in a conservative household and attended a church which instilled in him a deep disgust for homosexuals, as well as the judgment that being homosexual is immoral. If John later comes to think that his original judgment

about homosexuality is wrong, it appears as though Nussbaum is committed to the view that his disgust must also change. Yet it is easy to imagine that John might still be disgusted by homosexuals, though he no longer believes such an emotion is justified. How might Nussbaum explain situations where a person's emotions do not line up with his judgment?

Nussbaum's view appears to be that John would simply hold two contradictory judgments. One judgment (which is constituted by his disgust toward homosexuals) simply contradicts another more recent (and less emotional) judgment (Nussbaum 2001, 36). Here, though, Nussbaum's claim that judgments are sufficient for emotions presents another difficulty. For if John has a more recent (positive) judgment toward homosexuals, should he not have an emotion which corresponds with that judgment? After all, certain judgments are supposed to be sufficient for certain emotions.

Nussbaum may be able to respond in one of two ways. First, she might claim that John's more recent judgment about homosexuals does involve an emotion (perhaps respect). This would amount to saying that John is simply emotionally conflicted. He has two different, conflicting emotions about homosexuals, neither of which has gained the upper hand. Second, Nussbaum might claim that John's more recent judgment is a simple cognitive judgment, and does not have a corresponding emotion. This is allowed by her account, because while *certain* judgments are sufficient for *certain* emotions, it does not rule out the possibility that some judgments have no corresponding emotion (judgments about pedestrian facts like "when on the Washington, DC Metro escalators, I

should stand on the right and pass on the left" seem to plausibly have no corresponding emotion, for instance).

It is not clear that either of these responses match our intuitions about value judgments. First, there seems to be something important about what a person explicitly avows about his beliefs. If when asked, John says that he no longer believes that homosexuals are immoral, it seems strange to insist that John *really* also has the opposite belief if he admits that he still finds homosexuals disgusting. Surely the best access (from our perspective) to what he thinks he believes is what he avows, and in this case, he explicitly disavows the belief that Nussbaum would want to attribute to him.

Of course, Nussbaum cannot simply insist that John has a negative belief toward homosexuals because it is a result of her theory, because this would be begging the question. She might instead say that we are able to infer what John believes from his actions. Perhaps John goes out of his way to avoid a gay pride parade; this might allow us to infer that he thinks that homosexuals are immoral. But is this explanation for his behavior (which he explicitly disavows) any more reasonable than the alternative explanation (which he explicitly avows), namely, that he is merely disgusted by homosexuals but does not have a belief that they are immoral? Apart from question-begging, there seems to be no independent reason for preferring this explanation of his behavior. If anything, in such cases it seems reasonable to believe what a person explicitly avows about his beliefs.

This is not the only potential problem for Nussbaum's theory of emotions. In fact, there is a far more general criticism available. Andrea Scarantino has argued that

constitutive cognitivism suffers from what he calls the "Problem of Multiple Components": for any given emotion, there are various components that we associate with the emotion. Scarantino gives the example of anger:

> [W]e can distinguish in the complex phenomenon that is anger several parts: an evaluative component (e.g., evaluating being denied tenure as a slight), a physiological component (e.g., increased heart rate and blood pressure), a phenomenological component (e.g., an unpleasant feeling), an expressive component (e.g., fixed stare, loud voice, erected body), a physical action component (e.g., insulting, storming out of the room), and a mental action component (e.g., focusing attention, planning an appeal, remembering previous slights). (Scarantino 2010, 748)

Scarantino believes that such a group of heterogeneous components forces the constitutive cognitivist into a dilemma. One possibility is that her theory would be "falsified on a grand scale" because it fails to explain instances of fear which do not involve conscious awareness, or because it cannot "capture all nonevaluative components of emotions" like physiological changes (Scarantino 2010, 749).

There is one possible response: the constitutive cognitivist could simply dig in her heels and say that when she speaks about judgment, she *means* all of these different components. In this case, her use of the word "judgment" is simply a placeholder for this heterogeneous group. But Scarantino thinks this will not do for two reasons. First, such a solution simply renders the thesis of constitutive cognitivism trivially true. Of course judgments *are* emotions if by "judgment" we just mean all of the changes an emotion elicits in an organism. Such a thesis may be true, but will not explain anything. Second, such a thesis would no longer be falsifiable. There seems to be no empirical evidence which would disprove such an identification of judgment with emotion (Scarantino 2010, 749).

Scarantino thinks the "digging in her heels" move is quite common in constitutive cognitivist literature, and he calls it the "Elastic Strategy." He defines the move somewhat formally:

> The anti-cognitivist brings up a property P that emotions have … and judgments conservatively understood lack, or a property Q that judgments conservatively understood have … and emotions lack. On this basis, the anti-cognitivist concludes that emotions are not judgments, and that cognitivism is therefore falsified as a general theory of emotions. The Elastic Strategy consists of arguing that, under a liberal understanding of judgment, judgments also have property P … or lack property Q. (Scarantino 2010, 745).

Scarantino thinks that constitutive cognitivism will have to employ such a strategy against basically every counterexample provided by anti-cognitivists. No matter what the counterexample, this move will lead to the dilemma Scarantino posed above. Either constitutive cognitivism will be falsified, or it will adapt to incorporate the counterexample and thus become trivially true (nonexplanatory) or unfalsifiable.

I contend that this is a telling criticism of Nussbaum's view. Whatever the correct view of the interaction between judgments and emotions, that view must explain both judgments *and* emotions, and must not do so at the price of becoming unfalsifiable. There is another thesis available in the literature that promises to do precisely this: what Scarantino calls "etiological cognitivism." On this view, certain judgments are not themselves certain emotions. Instead, certain judgments *cause* certain emotions. Stated formally, this account says that "for all E, emotion E is caused by the appraisal that the formal object of E is instantiated (by some material object O)" (Scarantino 2010, 750).

This kind of account is more thoroughly fleshed out by Richard Lazarus, who believes that each emotion is not only defined by a specific meaning, but that these

(cognitive) meanings are the cause of certain emotions (Lazarus 1991, 39). Here is a quick sketch of the causal theory Lazarus employs: a person receives sensory input, (perhaps subconsciously) appraises the information generated from that sensory input. If her appraisal indicates that something in the situation instantiates the core relational theme[8] of some emotion, her appraisal will cause that emotion.

This is best explained with a (perhaps simplistic) example. A woman, Carol, is walking out of the supermarket when a stranger bolts by and steals her purse. Carol receives sensory input, which might include the physical features of the man, his jerking the purse out of her hand, etc. She then appraises (judges) the information generated from her sensory input. This appraisal indicates that he has stolen her purse. From here, the account would say that this appraisal causes the emotion of anger. The appraisal is the cause of the anger because it indicates to the agent that the situation in question (the man stealing her purse) instantiates the core relational theme corresponding to anger.[9]

Lazarus further breaks down the concept of appraisal at work here. There are two important parts of appraisal, both of which can be found in the example above. There is primary appraisal, which is "directed at the establishment of the significance or meaning of the event to the organism" (Lazarus 2001, 23). This part of appraisal can be further subdivided into three specific categories: goal-relevance, goal-congruence, and type-ego involvement. Goal relevance here is "whether an encounter is viewed by a person as relevant to well being," and Lazarus claims that there will be no emotion if

[8] Instead of speaking about the "formal object" of an emotion, Lazarus refers to "core relational themes." As Scarantino points out, these terms are interchangeable (Scarantino, 733).
[9] Drawing on Kenny (1963), Scarantino says that the formal object of anger is "A demeaning offense against me and mine" (Scarantino, 733).

there is no goal at stake (Lazarus 2001, 55). Goal (in)congruence is the agent's determination of "whether the conditions of an encounter facilitate or thwart what the person wants" (Lazarus 2001, 56). Finally, type-ego involvement includes "self or social esteem, moral values, ego ideals, commitment to certain meanings and ideas, the well-being of other persons, and life goals" (Lazarus 2001, 57). Then there is secondary appraisal, which is "directed at the assessment of the ability of the organism to cope with the consequences of the event" (ibid.). The features of secondary appraisal are blame, coping potential, and future expectancy (Lazarus 1991, 226). Lazarus provides the following table for anger:

> *Primary Appraisal Components*
> 1. If there is goal relevance, then any emotion is possible, including anger. If not, no emotion.
> 2. If there is goal incongruence, then only negative emotions are possible, including anger.
> 3. If the type of ego-involvement engaged is to preserve or enhance the self- or social-esteem aspect of one's ego-identity, then the emotion possibilities include anger, anxiety, and pride.
>
> *Secondary Appraisal Components*
> 4. If there is blame, which derives from the knowledge that someone is accountable for the harmful actions, and they could have been controlled, then anger occurs. If the blame is to another, the anger is directed externally; if to oneself, the anger is directed internally.
> 5. If coping potential favors attack as viable, then anger is facilitated.
> 6. If future expectancy is positive about the environmental response to attack, then anger is facilitated. (ibid.)

Lazarus notes that (1)-(4) are to be considered appraisal components necessary and sufficient for anger.

Let us see more specifically how Lazarus would evaluate our example. When Carol sees the stranger, she (1) appraises his actions as goal-relevant (he has run by *her*

and taken *her* purse). She also (2) appraises his presence as goal-incongruent (what he has done will cost her money). And she (3) appraises what he is doing as an offense against her (he has taken her purse unfairly). These components comprise her primary appraisal of the situation. Her secondary appraisal of the situation includes the next three components. (4) She sees the man as blameworthy in this situation; he quite evidently did not steal her purse on accident. According to Lazarus, the presence of (1)-(4) in this example will cause anger. He also notes that (5) and (6) can facilitate anger if aligned correctly. So, for example (5) if Carol thinks she can run to catch up with the man, her anger will be facilitated, whereas if he is much faster than her, she may be more resigned. The same is true for future expectancy (6) if she sees that a police officer has noticed the theft and given chase, this will facilitate her anger, whereas if she has no way of catching him, this will tend to make her more resigned.

What is important about this account is that these features of primary and secondary appraisal are supposed to be necessary and sufficient *cause*s of emotions. Though Lazarus indicates that there are different requirements for each emotion's core relational theme, he believes that we can spell out these requirements. In the case of anger, when there is a combination of goal relevance, goal incongruence, involvement of self or social esteem, and blame, anger *will* be produced.

This picture has been subject to one very important criticism: it seems unlikely that some emotions, such as fear, need to be so cognitively mediated. There seems to be a large difference between the cat that fears and runs from my dog and my fear about failing to get through graduate school. We would be more hesitant to ascribe a complex

cognitive appraisal to the cat's experience than to mine, yet we would still label both *fear*. This line of critique has been developed far more fully in Zajonc (1980).

There is a crucial presupposition in this criticism, however: it identifies "cognitive" with "conscious and well thought out." The etiological cognitivist, as Scarantino points out, can easily avoid the criticism by claiming that cognitive appraisals (judgments) may be relatively primitive and unconscious. Rather than assuming that all emotions must be caused by a conscious process, we might rightly see cognitive appraisals as lying on a continuum: this would put the cat fearing my dog's approach on the primitive end of the continuum and my fear of failing graduate school on the more sophisticated end.

Scarantino is somewhat skeptical of this approach, however. This is largely because of what he describes as the Problem of Levels of Appraisal. That is, if we treat appraisal as the cause of emotions, but admit that the same emotions can be caused by both sophisticated and primitive appraisals, "the class of appraisals becomes very heterogeneous" and this "glosses over the important differences between lower and higher levels of information processing in terms of speed and automaticity, neural pathways, evolutionary history, insulation from higher thought processes, and so on" (Scarantino 2010, 755). Thus etiological cognitivism runs the risk of not being very explanatory.

I do not believe this is a particularly telling criticism of etiological cognitivism. Even Scarantino admits that there has been some work done on distinguishing these levels (he points to the work of Leventhal and Scherer (1987), van Reekum and Scherer

(1997), and Teasdale (1999)). It appears as though etiological cognitivism could be sufficiently explanatory as long as these different levels of appraisal were mapped out in adequate detail. From here I will adopt etiological cognitivism: judgments cause emotions, and judgments can be arrayed on a continuum going from relatively primitive judgments to more sophisticated, cognitively complex judgments. In the next section I will point to evidence that specifically *moral* judgments are on the primitive end of this spectrum.

### 3.2 The SI Model: Intuitions Are Primary

At this point, it is important to see where we stand. There seem to be good reasons to resist the idea that judgments constitute emotions. As Scarantino has shown quite convincingly, identifying emotions with cognitive appraisals is a project that fails to explain much about both emotions and judgments. Moreover, the account is (for reasons argued previously) either trivially true or unfalsifiable. As a result, I raised another possibility about the connection between emotions and judgments: judgments *cause* emotions. In this section I will draw on the work of Jonathan Haidt to argue that moral judgments are typically *intuitive* judgments. I will use the term "intuition" to refer to any given intuitive moral judgment and the term "reasoned judgment" to refer to any given reasoned moral judgment. These can be seen as strict analogues of what Scarantino calls "primitive appraisals" and "sophisticated appraisals," respectively, and are not meant to connote what philosophers have historically attributed to either term. As we will see in the section that follows this one, this argument causes serious problems for coherentist accounts of moral realism.

First it is important to get clear on what exactly distinguishes intuitions from reasoned judgments. First, intuition generally occurs quickly, whereas reasoning occurs more slowly. Second, only the product of intuition is accessible to consciousness, whereas in the case of reasoning, both the product and the process are so accessible. Third, intuition seems more or less "mandatory," in that the product of intuition arrives in consciousness without any explicit decision on the part of the agent, whereas reasoning is not mandatory; if I decide not to reason about something, I do not get any product of reasoning at all. As Haidt puts it, "intuition occurs quickly, effortlessly, and automatically, such that the outcome but not the process is accessible to consciousness, whereas reasoning occurs more slowly, requires some effort, and involves at least some steps that are accessible to consciousness" (Haidt 2001, 818).

There are two main accounts available for how moral judgments might operate under etiological cognitivism. What we might call the rationalist model claims that moral judgments are typically reasoned judgments, and that these judgments cause moral emotions. By contrast, the SI model holds that moral judgments are typically intuitions, and that these cause moral emotions. Here I will turn to an example to illustrate the difference, drawing again on Lazarus' account of etiological cognitivism.

Suppose Elizabeth sees two men holding hands. She processes this visual input and forms a judgment about the information it contains. Her primary appraisal of the situation must then indicate whether what she sees is goal-relevant, congruent with her goals, and ego-involved. Here, we can see that how she appraises the situation will depend on various beliefs or dispositional feelings she might have. If Elizabeth is a

social liberal, she may not have any negative beliefs or feelings toward two men holding hands. If this is the case, she may appraise the situation as irrelevant to her goals and the two other features of primary appraisal need not even be invoked. Something which is irrelevant to one's goals is neither congruent nor incongruent with them, and further, will not involve the person's ego. What this amounts to saying is that Elizabeth would form no specific emotion as a result of seeing the two men holding hands. Moreover, it seems perfectly consistent to say that she might not form any judgment whatsoever about her experience.

Now take another woman, Sarah, who is a social conservative and deeply disgusted by any overt display of homosexuality. Lazarus provides the following table for disgust:

> Primary Appraisal Components
> 1. If there is goal relevance, then any emotion is possible, including disgust.
> 2. If there is goal incongruence, then only negative emotions are possible, including disgust.
> 3. If any of the six types of ego-involvement is at risk of being contaminated by a "poisonous idea," then disgust will occur.
> [Note:] No secondary appraisal components are essential. (Lazarus, 261)

This account will help us to explain why Sarah experiences the emotion she does. When she sees the same two men holding hands, she appraises it much differently than Elizabeth did. Her primary appraisal of her sensory information is that it is relevant to her goals (avoiding homosexuals) and incongruent with her goals (she is not now avoiding homosexuals). For the last component of Sarah's primary appraisal, Lazarus provides a list of six different kinds of ego-involvement: self- and social esteem, moral values, ego-ideals, meanings and ideas, other persons and their well-being, and life goals

(Lazarus 2001, 58). There are a number of possibilities here, but we should see that if any of these ego-involvement categories is "at risk of being contaminated by [metaphorically speaking] a 'poisonous idea,' then disgust will occur" (Lazarus 1991, 261). Virtually any of these types of ego-involvement could be present, though we should resist explaining her emotion on the basis of moral values, as that would make the forthcoming account look circular. The point is that, according to etiological cognitivism, her disgust is caused by a certain kind of judgment. On this account, her judgment might very well be intuitive. Sarah need not actively think about what her moral values indicate she should feel; she might just be the kind of person who does not like to be around homosexuals.

Now, if we ask Sarah why she formed the judgment she does, and ask Elizabeth why she did not form a judgment at all, each would typically be able to provide a reason. Perhaps Elizabeth would say she does not care because two men holding hands does not harm her in any way. Perhaps Sarah would say that she is offended (disgusted) because two men holding hands (at least in most Western culture) indicates a romantic interest, and she believes such a romantic interest has been forbidden by God. Each woman would thus be both explaining and justifying her judgment (or lack thereof).

Now, the crucial question for the etiological cognitivist is this: what should we consider to be the *cause* of the judgment (or lack of judgment) each woman makes? Should we accept what each woman says when she responds (namely, that each has a belief which entails either a judgment or no judgment), or should we say that the actual cause is each woman's *intuitive* appraisal of the situation? This question hinges on

whether we have good evidence to suppose that each woman arrived at her judgment or lack thereof through a mental process accessible to consciousness, or whether the product of the judgment arrived in her consciousness spontaneously. This will differentiate whether her judgment of the situation was a reasoned or intuition.

The rationalist model takes it that we have good evidence to believe the former. The SI model takes it that we have good evidence for the latter. In what follows, I will relate some of the evidence in favor of the SI model. I will argue that the cause of at least *most* moral judgments is not the moral beliefs one consciously draws up to defend one's judgment, but rather intuitive judgments which are not consciously accessible. As Haidt argues, "when faced with a social demand for a verbal justification [of one's judgment], one becomes a lawyer trying to build a case rather than a judge searching for the truth" (Haidt 2001, 814).

Before proceeding, we should be careful to distinguish three possibilities for which kinds of moral judgments there might be: (1) moral judgments are always sophisticated judgments, (2) moral judgments are always intuitive judgments, and (3) both kinds of judgments exist. As one might be able to see, both (1) and (2) are quite strong theses.

It is not immediately clear whether Haidt holds (2) or (3). He does claim that "The [SI] model proposes that moral judgments appear in consciousness automatically and effortlessly as the result of moral intuitions" (Haidt 2001, 818). Haidt often speaks of subjects' justification for their moral judgment in a situation as "post hoc." Indeed, this is a core tenet of the SI model (Haidt 2001, 822). The idea is not that people never

produce reasoned judgments, but that when subjects do make them, they are simply an attempt to dress up an intuition. One could construe this as the claim that reasoned judgments do not exist, but are rather just *rationalizations*. Alternatively, we could see Haidt as recognizing the existence of reasoned judgments, but thinking they play a much more restricted role than the rationalist thinks they do. I think the latter interpretation is better, as calling reasoned judgments "rationalizations" runs the risk of begging the question on the normative issues I will mention in the next section.

It is also important to distinctly characterize the restricted role of reasoned judgments in the SI model. To this end, we should distinguish two different types of reasoned judgments. An *ultimately* reasoned judgment is a moral judgment which is not caused or (subjectively) justified by any intuition. A *proximately* reasoned judgment is the opposite: a moral judgment which is caused or (subjectively) justified by an intuition, though the causal or justificatory chain may be long and complex.[10] With this distinction in hand, Haidt's position can be described thusly: reasoned judgments exist, but they are always proximately reasoned judgments. This makes the rationalist alternative particularly vivid as well: reasoned judgments exist, and at least some of them are ultimately reasoned judgments.

There is a normative dimension involved here. That is, which kind of judgment should we consider justifying? There seem to be a number of positions one could hold: (A) Intuitions never justify other moral beliefs. (B) Only intuitions justify other moral

[10] In using subjective justification here because I mean to emphasize that the definition of "ultimately" and "proximately" reasoned moral judgments does not depend on the moral judgment's being *objectively* justified. One could imagine ultimately reasoned moral judgments which are in fact objectively unjustified, though the person doing the judging *thinks* they are justified.

beliefs. Both of these possibilities might (but do not necessarily) represent forms of foundationalism. Here there is a third option as well: (C) some mix of (moral) intuitions and reasoned (moral) judgments is justifying, as long as this mix is coherent. This represents reflective equilibrium, or coherentism. I mention this question in order to shelve it, as it will become apropos only in the next section.

Now, on to the evidence. The SI model's main hypothesis, that reasoned judgments are always proximately reasoned judgments, can be experimentally verified by means of scenarios which elicit moral emotions (and thus, under etiological cognitivism, moral judgments) but give subjects little with which to consciously reason. As a contrast case, first consider a situation where a subject would have *much* to reason with. If experimenters pose a scenario to a subject in which someone steals clothes from a department store, the subject will find it easy to explain and justify her anger about the action with reference to the harms that the theft causes. She might say that one should not steal because it harms the store, causes prices to go up for its patrons, etc. This is a situation where her emotional judgment is strongly supported by reasons she gives, and that fact makes a rationalist interpretation plausible. That is, the rationalist would say that she appraises the hypothetical by consciously noting that the situation is relevant, goal incongruent, ego-involved, and includes blame. All of these elements can be (perhaps in less technical terms) avowed by the subject as reasons for her moral judgment and the resultant moral emotion (anger). These kinds of examples have traditionally led philosophers and psychologists to suppose that for moral judgments

generally, the reasons one consciously gives for one's judgment are the causes of that judgment (and thus also the causes of one's moral emotion).

Yet we now have strong evidence that this generalization does not hold. This is because when subjects in a similar experiment are asked to judge a scenario involving putatively morally objectionable actions, *but these actions have been stripped of any possible harmfulness*, we find that subjects do not change their moral judgments, and the belief that these actions are morally wrong endures. For example, one experiment asked questions which involved eating one's already dead pet dog, cleaning one's toilet with the national flag, or eating a chicken carcass one has just used for masturbation (Haidt et al. 1993, 617). Even though subjects could not explain how any harm was involved, they still judged that the actions described in the experiment were wrong. If the conscious reasons subjects give were really the cause of their moral judgments, it would be difficult to explain the tenacity with which these subjects hold judgments in these cases. One would think that once one's reasons for holding a judgment had been defeated, if those reasons really were the cause of one's moral judgment, one would (perhaps after some time) reverse such a judgment. As Haidt's experiments show, this is definitely not the case.

In another experiment, Haidt aimed to get at exactly how difficult these intuitive moral judgments are to reverse. This experiment involved five different stories, two of which I will focus on here: the cannibalism story and the incest story. The cannibalism story involves a medical school student who cuts off a piece of a human cadaver, cooks it thoroughly (to eliminate the risk of disease), and eats it. The incest story involves a

brother and sister who decide to have sex while on vacation in France. Both use a form of birth control, they only do it once, and "[t]hey keep that night as a special secret between them, which makes them feel even closer to each other" (Haidt et al. 2000, 16). In this experiment, Haidt's experimenters played "devil's advocate," and questioned the reasons that subjects gave in response to these stories. In both stories, experimenters pointed out to negatively judging subjects that there was no harm done in the scenario.

The experiment's results were surprising. For the cannibalism and incest scenarios, subjects *very seldom* changed their judgments when their arguments were (ostensibly) defeated by the experimenters. For the cannibalism scenario, only thirteen percent judged that the action in question was morally permissible, and this number only rose to twenty eight percent after the experimenters rebutted the arguments of those who judged negatively. This means that only seventeen percent of the subjects changed their minds. For the incest scenario, twenty percent judged that the action in question was morally wrong, and again after having their arguments rebutted, only seventeen percent of the overall participants changed their minds (Haidt et al, 2000, 14).

What this evidence indicates is that for at least some moral questions, individuals' moral judgments are highly intransigent. That is, even when told how their arguments fail, subjects still overwhelmingly hold to their judgment even in the absence of compelling *reasons* for holding it. This gives us some preliminary evidence that when there is a conflict between an intuition and a reasoned judgment, intuition generally wins out.

In these kinds of experiments Haidt reports a phenomenon he calls "moral dumbfounding": the tendency of subjects to "stutter, laugh, and express surprise at their inability to find supporting reasons" when standard harm-based objections to the behavior in question have been preempted. This phenomenon is especially prevalent in Haidt & Hersh (2001), which involves an experiment that examines the differences between liberals and conservatives with regard to sexual morality. Students (broadly self-classified as liberals, moderates, or conservatives) were presented with a number of scenarios, some of which involved homosexual anal sex, a woman masturbating while cuddling with her teddy bear, and consensual incest between a brother and sister where both used a form of birth control. Moral dumbfounding was extremely common (though not omnipresent) in these scenarios. In the scenario involving homosexual sex, sixty percent of subjects self-classified as conservative exhibited moral dumbfounding. Subjects self-classified as liberals demonstrated moral dumbfounding at the same rate (sixty percent) for a scenario in which "a man likes to masturbate while his dog willingly licks his owner's genitals and seems to enjoy it" (Haidt & Hersh 2001, p. 197). Remember what this means: when asked about these scenarios, a full sixty percent of these groups of subjects could not articulate any reason for their positions and (perhaps understandably, given the content) showed signs of discomfort, stuttering, and laughing. Moreover, for the former scenario, conservatives baldly stated their affective response as the justification for their view at a rate of forty percent, with liberals doing the same for the latter scenario at a rate of thirty eight percent (ibid., p. 209).

What these experiments are intended to show is that the reasoned judgments the subjects provided were entirely parasitic on their intuitions. To illustrate this fact, consider the following example where my reasoned judgment is *not* parasitic in this way. Suppose I tell you that I believe that the next train will come into the station at 12:15PM. You ask me why I think this is the case. I show you a schedule which indicates that time. But you notice that the schedule I have is outdated: the new schedule, which you have in your hand, indicates it will come in at 12:30 PM. You point out this fact, and realizing I have made a mistake, I change my belief. In this case I formed a reasoned judgment that the train would come in at 12:15 PM; I relied on the schedule I have to consciously infer that the train will arrive at that time. But upon being told that the schedule I used to infer this information is unreliable, I realize that my inference was incorrect and change my belief.

Now, if I were to acknowledge that the schedule was unreliable but still believed that I knew the train would come in at 12:15PM, this would show that my reasoned judgment was *not* in fact the cause of my belief. In such a case, you might ask me why I believe that the train will come in at that time. If I cannot provide any reason whatsoever, it appears that the only possible explanation for my belief is that I have formed some intuition that the train *must* come in at this time, despite all the facts at my disposal. In these cases it is usually right to say that, while I have no *reason* to believe what I do, I have a "hunch" or a "gut feeling" that it will.

This kind of gut feeling appears to be what is driving the subjects' responses in these experiments. When the reasons which purport to factor into their reasoned

judgment are preempted (much like you pointing out that my train schedule is unreliable), the subjects cling to the judgment anyway. The reasoned judgments that subjects make in these cases do not independently support the subjects' intuitive judgments.

What is not entirely clear is whether the evidence adduced above licenses us in believing that *all* reasoned judgments are proximate in this sense. It may be that for these strange cases people rely on intuitions, but that in the normal case, the rationalist account is correct. In line with this reasoning, I should note one interesting wrinkle in the experiment on sexual morality just described (Haidt & Hersh 2001), which might cause problems for the SI model. Haidt found that conservatives were generally much more likely to become morally dumbfounded than liberals. The difference, Haidt thinks, is that liberals "often separated their own emotional reactions from their moral judgments" (ibid., p. 195). Haidt quotes one liberal subject in response to a question on the permissibility of gay sex:

> I'd have to say it would make me feel weird. That's my first instinct. Yeah, I mean there's a weird kind of thing going on in my head there. I feel like there's a way I'm going to act and a way to act based on this preconceived ideology I have about sex being okay between people, no matter what sex they are. You know, instantly a warning bell comes up, and that bugs me. I quickly discard it; I mean I recognize it and then discard it because it makes me feel silly. (ibid, p. 195)

This might lead one to believe that in this experiment at least some liberals were making ultimately reasoned judgments that independently *contradicted* their intuition, and that they favored the reasoned judgment over the intuitive alternative. Engaging in a bit of warranted speculation, the rationalist might say that this liberal respondent came to the intuitive judgment that homosexual sex was wrong (hence, on etiological cognitivist

grounds, formed the emotion of disgust), but overrode this initial intuition and instead identified himself with the reasoned judgment that gay sex is permissible. Perhaps the liberal's reasoned judgment could be put into a neat, conscious inference pattern like this:

Premise 1: Disgust is not a justified basis for a moral judgment.

Premise 2: I feel disgust at the idea of gay sex, but by premise 1, this feeling is irrelevant.

Premise 3: I believe in the harm principle: an action is only wrong when it harms someone non-consensually.

Premise 4: I have no reason to believe gay sex will harm anyone non-consensually.

Conclusion: By premise 3 and 4, gay sex is morally permissible.

It seems entirely plausible that such a model could be part of a reasoned judgment, and that the liberal in Haidt's study reasoned something like this. Of course, if Haidt were to deny this (which I doubt), I could point to another person who *does* reason like this: me.

How would a proponent of the SI model respond? First, he might say that in this and other experiments there is some bias in favor of providing reasoned instead of intuitive judgments. In all of Haidt's experiments, there is some pressure on respondents to provide reasons and not to only rely on how the scenario *feels* to them. Thus respondents who would have given an intuition are instead pressured to provide a reasoned judgment.

Second, he might note that Haidt's experiments do not support the general conclusion that liberals *always* (or even often) use reasoned judgments to override their intuitions. Consider again one of the scenarios posed to liberal and conservative subjects: "A 25-year-old man likes to masturbate while his dog willingly licks his owner's genitals and seems to enjoy it" (Haidt & Hersh 2001, 197). Earlier I mentioned moral dumbfounding in connection with this scenario, but just as salient is the fact that both conservatives and liberals state their affective response as a reason for their judgment at nearly the same rates (forty percent for conservatives, thirty-three percent for liberals) (Haidt & Hersh 2001, 209). Liberals are clearly not overriding their intuitive judgment in this case. These issues aside, what is most telling is that while liberals were indeed significantly less likely to state their affective response as the reason for their judgment (ibid.), this only occurred in response to scenarios where liberals had significantly less negative affect (the two scenarios where this result obtained were one involving gay male anal sex and another involving lesbian oral sex) (ibid., 203). Despite the one example cited above, Haidt's experiments provide evidence that liberals generally do not condemn certain behaviors because they feel no negative affect, not because they are substituting an intuition for a reasoned judgment.

These responses are not quite good enough though; though dumbfounding is frequent despite a bias in the experiments in favor of independent, ultimately reasoned judgments, the fact that subjects can make ultimately reasoned judgments contradicts the main hypothesis of the SI model; the fact that liberals do not in general make ultimately reasoned judgments is not enough to save the SI model from this counterexample.

The proponent of the SI model must do more than either of these responses allow. He must argue that while the liberal's judgment does constitute an instance of a reasoned judgment overriding an intuition, the overriding judgment is necessarily a *proximately* reasoned judgment. That is, somewhere along the line of explanation and justification, the liberal's reasoned judgment must bottom out in an intuition. Otherwise the thesis of the SI model is falsified.

This line of argument seems to be implicit in a good deal of the SI model literature. For example, Haidt explains the aforementioned liberal's response this way: "because his political culture (liberalism) has a preconceived ideology that sex is entirely a matter of personal choice, he feels compelled to override his intuition" (Haidt & Hersh 2001, 212). Haidt sees this political culture as a contingent factor influencing which intuitions a person is liable to accept and which he is liable to override with a reasoned judgment. Uninterpreted, it is easy to see that admitting this would still pose a problem for the SI model. This is because one's political culture often has many reasoned judgments implicit in it (the harm principle being one in the case of the aforementioned liberal). If Haidt is willing to admit that these kinds of reasoned judgments occur, he must find a way to claim that it is because those reasoned judgments are themselves the results of intuitions somewhere down the line.

We see now that the viability of the SI model depends on providing a plausible story for the genesis of reasoned judgments. I do not contend that this matter has been settled. However, there is one body of research which is promising for the SI model's proponent: the CAD triad hypothesis. This hypothesis claims that "[three] emotions—

contempt, anger, and disgust—are typically elicited, across cultures, by violations of [three] moral codes" (Rozin et al. 1999, 574). These moral codes can be understood as families of moral judgments. On this hypothesis, one of three main moral emotions is elicited as a result of a given violation: contempt, anger, and disgust. Rozin et al. label these three codes community, autonomy, and divinity. The first emotion, contempt, is associated with violations of the ethics of community. This emotion is triggered when an individual fails to conform to social rules regarding rank and hierarchy. The second emotion, anger, is associated with violations of the ethics of autonomy. This emotion arises as a response to "insults, transgressions, and rights violations against the self or those close to the self" (Rozin et al. 1999, 575). The third emotion, disgust, is associated with violations of the ethics of divinity. It is triggered when an individual violates norms established to protect the soul or world against pollution or degradation (ibid.).

These three families of moral judgments are further defined as follows:

1. [The ethics of *Autonomy*] Individual freedom/rights violations. In these cases an action is wrong because it directly hurts another person, or infringes upon his/her rights or freedoms as an individual. To decide if an action is wrong, you think about things like harm, rights, justice, freedom, fairness, individualism, and the importance of individual choice and liberty.

2. [The ethics of *Community*] In these cases an action is wrong because a person fails to carry out his or her duties within a community, or to the social hierarchy within the community. To decide if an action is wrong, you think about things like duty, role-obligation, respect for authority, loyalty, group honor, interdependence, and the preservation of the community.

3. [The ethics of *Divinity*] Divinity/purity violations. In these cases a person disrespects the sacredness of God, or causes impurity or degradation to himself/herself, or to others. To decide if an action is wrong, you think about things like sin, the natural order of things, sanctity, and the protection of the soul or the world from degradation and spiritual defilement. (Rozin et al. 1999, 575-576)

First let me make a terminological point. There has been some question about whether "the ethics of divinity" is the proper name for a family of moral judgments which does not always invoke a deity. Following Prinz, I will henceforth refer to "the ethics of divinity" as "the ethics of the natural order" (Prinz 2007, 73).

In tandem with these three basic moral emotions, three other emotions form the basis of moral responses to condemnation in the three CAD areas: shame, embarrassment, and guilt. These three emotions form the SEG triad, and each serve to "motivate the individual to want to fit in, to behave in a culturally acceptable fashion, and to avoid harming people" (Rozin et al. 1999, 574). These emotions appear to be responses to contempt, anger, or disgust on the part of others, but are not well correlated with any of these individual emotions. This element of the theory is not yet fully understood, as these emotions appear to be the subject of more cultural variation than the CAD triad. Work is still under way in an effort to taxonomize them.

Once again, the CAD triad hypothesis claims that these three moral emotions are typically elicited cross-culturally as a result of violations of one of the families of moral judgments. This hypothesis has so far held up to experimentation. Rozin et al. (1999) conducted an experiment in which American and Japanese participants were read different scenarios in which some rule violation occurs and asked which of the three emotions a hypothetical person who witnessed such a violation would express. Subjects responded in two ways. First, they indicated which facial expressions they expected the hypothetical person to display. Second, they described the emotion that the hypothetical witness would feel toward the violator, and were limited to three options: contempt,

anger, and disgust (and their Japanese translations). The results were impressive, though not perfect. The model predicted which facial expression subjects would choose in all three categories of rule violations with a high degree of accuracy. More respondents chose the predicted facial expression than chose either of the other two facial expressions combined, which far outstrips mere chance. The model also predicted which word the subjects would choose in response to each of the hypothetical rule violations, though with somewhat less accuracy than on the facial recognition task. This experiment lends at least prima facie credibility to the idea that different moral judgments are accompanied by different moral emotions, and that these emotions can be taxonomized effectively.

The CAD triad hypothesis represents a possible account of the genesis of reasoned judgments for the SI model. The basic idea is put quite well by Haidt:

> A child is born prepared to develop moral intuitions in all three ethics, but her local cultural environment generally stresses only one or two of the ethics. Intuitions within culturally supported ethics become sharper and more chronically accessible … whereas intuitions within unsupported ethics become weaker and less accessible. (Haidt 2001, 827)

On this theory, the three codes of the CAD triad are families of moral judgments which come naturally to human beings, but are stressed differently by different cultures.[11] If one culture stresses the ethics of the natural order, children growing up in that culture will be apt to express intuitions which conform to that code.

[11] Though Haidt appears to think that culture has the largest impact on the variation in which moral codes are emphasized, we need not think culture is the only factor, or, *contra* Haidt, even the most important factor in code variation. The important thing for the argument at hand is the existence of the variation, not the *source* of it, which is another question entirely.

The SI model's proponent must hold that reasoned judgments are always inferred from intuitions. To see how this would happen, consider two different children, Hector and Alyssa. Hector grows up in a culture which stresses the ethics of the natural order. He is taught that feeling disgust in a scenario is a justified basis for a negative moral judgment against whoever is involved in that scenario. Alyssa grows up in a culture which stresses the ethics of autonomy and downplays the ethics of the natural order. As a result, Alyssa is taught more or less the opposite of Hector: when she experiences disgust, she is taught that it is not a justified basis for a negative moral judgment. Now suppose these children grow up, and when they are adults, they are subjected to one of Haidt's experiments. When both individuals are given a scenario which evokes disgust in human beings *generally* (the existence of these scenarios is supported by the CAD triad hypothesis), *both* will form a negative intuition about the object of their disgust. Moreover, both will form a reasoned judgment as a result of this intuition. But because of the culture each was raised in, the reasoned judgment of each will be different! Hector will form a negative reasoned judgment which supports his intuition against the object of disgust. He might say that the scenario is offensive to God, or that it is unnatural, etc. Alyssa will have formed the same intuition as Hector, and this intuition will (as per etiological cognitivism) cause disgust. Because Alyssa's culture downplays the ethics of the natural order, however, she does not recognize disgust as a legitimate basis for a reasoned judgment. If the scenario involves some sort of harm, she may be able form a compatible reasoned judgment against the object of her disgust based on the ethics of autonomy. But if the scenario does not involve harm, she will have to form a

reasoned judgment that contradicts her intuition, and in all likelihood, will hold this reasoned judgment to be overriding.

Of course, the data in Haidt's experiments so far indicates that most people will simply judge the object of their disgust negatively, even when their reasoned judgment is obviously unsatisfactory. That is, most people will attempt to find harm in scenarios where harmfulness has been preemptively excluded. When the experimenters calmly point out that their arguments fail, most subjects fall back on their negative intuition about the subject.

Haidt also mentioned that we should expect certain intuitions from certain groups (e.g. liberals) to be weaker than the intuitions of others (e.g. conservatives), largely due to cultural influence. This expectation is supported in the evidence, where we see liberals expressing less affective response than conservatives (though the difference was not very large for some scenarios) (Haidt & Hersh 2001, 211). In keeping with etiological cognitivism, we can say that a weaker intuition causes a weaker emotional response. Some liberals had a weak enough intuition that they formed little emotional response at all and thus had no problem giving a reasoned judgment on the basis of the ethics of autonomy. Other liberals had a stronger intuition (and a stronger affective response), though because of their cultural background, they still relied on the ethics of autonomy to override their intuition.

This genesis story explains how the liberal in Haidt's experiments can form a reasoned judgment which overrides his intuition without forcing the SI model to admit that any reasoned judgments are *ultimately* reasoned judgments. When the liberal

overrides his intuition, the reasoned judgment he overrides it with is itself reducible to a set of intuitions within a different code of ethics (in this case, the ethics of autonomy). For when we ask why the overriding liberal identifies with his reasoned judgment instead of his intuition, it can only be because he accepts all the premises in an explicit reasoning pattern modeled above). From here we need only ask: why does he accept the premises? To restate, here is the argument that I proposed the liberal might be using:

Premise 1: Disgust is not a justified basis for a moral judgment.

Premise 2: I feel disgust at the idea of gay sex, but by premise 1, this feeling is irrelevant.

Premise 3: I believe in the harm principle: an action is only wrong when it harms someone non-consensually.

Premise 4: I have no reason to believe gay sex will harm anyone non-consensually.

Conclusion: By premise 3 and 4, gay sex is morally permissible.

Premises 2 and 4 are comprised of factual information about the scenario at hand. Premise 1 is caused by his immersion from early childhood in a culture which discounts the ethics of the natural order. And premise 4 is a combination of that immersion and of other intuitions in scenarios or experiences which involve harm. Therefore the liberal's reasoned judgment is ultimately caused by his culture and other intuitions. If this genesis story is correct, the liberal does not constitute a counterexample to the general thesis of the SI model: that all reasoned judgments are proximate reasoned judgments.

Let me sum up this section in somewhat more ordinary language. What the SI model amounts to is the claim that the ultimate cause of all moral judgments is intuition, not moral reasoning. Moral reasoning, where it exists, is always causally constrained by which types of moral intuitions (and their accompanying emotions) a culture happens to emphasize (or reject) as a sound basis for moral reasoning. We have some evidence that (for humans generally) a few moral emotions (contempt, anger, and disgust) are elicited by certain regular features of situations. Intuitions tend to converge in these areas, though there exist large variations both in the features cultures see as salient in these situations and in the types of intuitions cultures find acceptable in forming reasoned judgments about these situations. It appears that these cultural variations in moral intuitions are the root of moral disagreement generally, and the SI model combined with the CAD triad indicates that this root is buried very deep. If moral reasoning really is always just caused by moral intuition and culture, then this has serious implications for theories of moral epistemology which aim at convergence on moral truth *through* moral reasoning. In the next section, I will explore these implications.

### 3.3 Implications for Normative Theories

Earlier I mentioned (in passing) that there is a normative dimension in play here. That is, there are a number of positions one might hold with respect to which kinds of moral judgments are justifying. With the evidence for the SI model on the table, we are now in a position to evaluate what this evidence implies for such normative accounts. Recall the three options mentioned in the previous section: (A) intuitions never justify other moral beliefs, (B) only intuitions justify other moral beliefs, and (C) some mix of

moral intuitions and reasoned (moral) judgments is justifying. I mentioned that (A) and (B) might be plausibly considered versions of foundationalism. Though my argument is chiefly aimed at (C), it is worth taking a moment to examine why the evidence for the SI model also indicts normative accounts like (A) and (B). It is important to note that my argument here only applies to *objective* forms of foundationalism and coherentism which claim that there is one set of moral standards which apply to all individuals. Subjective forms of these theories, which admit of various forms of relativism, are not affected by the arguments that follow.

First, let us consider (A). An objective foundationalist who holds this position is committed to denying that *any* of the codes of ethics or the families of intuitions which these codes embody is a justifying basis for one's other moral beliefs. This is easy enough to see, because such a foundationalist denies this justifying role to any intuition whatsoever. This kind of foundationalist would say that whatever one's moral intuitions, one should always override them with a reasoned judgment. Sometimes our reasoned judgments and intuitions are in sync, and sometimes they are not, but when there is a conflict, we should always prefer the reasoned judgment.

The problem for this account is that I have given (what I consider to be) compelling evidence that reasoned moral judgments are always only proximately reasoned. That is, whenever one has a reasoned judgment, it is ultimately derived from the intuitions one's culture finds acceptable. If this is the case, it seems that there is no objective basis for preferring one's reasoned judgments over one's intuitions. Individuals from different cultures will provide different reasoned judgments, and we

have no objective means of determining which of these judgments to prefer. If this is the case, we are not capable of this kind of objective knowledge.

Next, consider (B). This would be a form of intuitionism. Any objective version of this account seems to be even more transparently undermined by the SI model. This is because there are significant differences in the intuitions of individuals from different cultural backgrounds. If only intuitions justify other moral beliefs, it seems that this form of objective foundationalism must hold that one set of these intuitions is the *correct set*. But how could we determine which set is correct? Obviously one cannot simply rely on one's own culture in determining this set of correct intuitions, because one very quickly runs the risk of ethnocentrism. But even more important is the fact that there seems to be no obvious and objective method of determining which code of ethics (and which family of intuitions) to emphasize or deemphasize. Thus if a form of foundationalism asserts that only intuitions ultimately justify reasoned judgments, it seems that this form of foundationalism must also necessarily be subjective.

Now, what about (C), which claims that some mix of intuitions and reasoned moral judgments is justifying, as long as this mix is coherent? There appear to be two problems for this kind of account. To get at these problems, we will need to understand two different theses that the coherentist might hold.

> *Descriptive Coherentism:* at least one individual actually can employ coherentism in her efforts to obtain subjective justification for her system of morality.

*Normative Coherentism*: Coherentism is how one *ought to* justify one's system of morality.

It might not be immediately apparent why this distinction is necessary. After all, we are talking about whether the SI model has implications for *normative* theories like coherentism, not whether individuals actually do use coherentist reasoning to get subjective justification. In what follows, I will argue that the evidence for the SI model puts *Descriptive Coherentism* in doubt, and that this in turn throws *Normative Coherentism* into doubt as well.

Consider what the evidence for the SI model indicates. First let us focus just on the cases where subjects cannot find reasoned judgments to support their response to a scenario, or where their reasoned judgments are quickly rebutted by the experimenters but their negative judgment persists. In these cases, we might rightly describe subjects' judgments as stubborn in the face of reasons. Indeed, pointing to these cases, Haidt argues that "[t]he refutation of … arguments does not cause people to change their minds; it only forces them to work harder to find replacement arguments" (Haidt & Hersh 2001, p. 218). If Haidt's experiments show anything, it is that in these cases, if we are to convince someone to change her position on a moral question, she will not be motivated to change that position *simply because* her reasons have been defeated. When a person refuses to change her position as a result of her reasons being defeated, we have good reason to believe that she has an intuition about the case at hand. As I have argued, in these cases her intuition is driving her judgment. It is these kinds of intuitions, which

drive moral judgment even in the absence of a reasoned judgment, which I will refer to as "intransigent."

It is not immediately obvious that the moral judgment a person identifies with is always caused by intransigent intuitions. Indeed, this is a strong thesis, and the rationalist might again point to the overriding liberal as a counterexample. However, here I need only point to the argument that all reasoned judgments are only proximately reasoned. The overriding liberal is not making an ultimately reasoned judgments. Rather, he is relying on an intransigent intuition as well; it is just not the intuition which is being tested in the experiment. Again, the idea here is that when the liberal overrides his intuitive disgust against the idea of gay sex, he is still relying on intuitions which would themselves be immune to defeat by any reasoned judgment. If we were to test the intuitions which he relies on in making his positive reasoned judgment about gay sex, I contend that we would find those intuitions just as intransigent as those of the conservative who does not override his intuition in this case at all. Of course, we should note that the overriding liberal still has (according to etiological cognitivism) a negative judgment about gay sex; it is just not the judgment he identifies with. The thesis I am advocating here is not that all intuitions are intransigent, but that all intuitions which cause the moral judgment one ultimately identifies with are intransigent. I will refer to this thesis as "the problem of intransigence."

This problem gives us reason to believe that *Descriptive Coherentism* is false. After all, if all the judgments one ultimately identifies with are based on intransigent intuitions, this immediately implies (by the *definition* of intransigent intuitions) that they

cannot be changed as a result of reasoned judgments. In this context, some intuitions appear to be (for lack of a better description) "anchored down" in one's psychology. They provide the bounds within which moral reasoning occurs in the first place. Thus it appears that the clean picture of coherentism from section two in which a person engages in reflective equilibrium with all doxastic options open to him is not psychologically accurate.

Of course, this does not mean that the moral judgments we identify with cannot be changed *at all.* Clearly culture (and perhaps a host of other sociological factors) influences which intuitions a person finds acceptable, and these factors have an effect over the course of one's lifetime. But these sociological factors influence a person at a level below that of conscious reasoning. Indeed, the problem of intransigence persists even when we appear to have evidence that someone has changed her mind as a result of an argument. As Haidt hypothesizes, "reasoned persuasion works not by providing logically compelling arguments but by triggering new affectively valenced intuitions in the listener" (Haidt 2001, 819). Reasoned argument may sometimes change an individual's mind because it directs her focus to another intuition that she has (a strategy Prinz calls "norm pitting" (Prinz 2007, 289)). However, if the SI model is correct, it cannot get her to drop an intuition in favor of a reasoned judgment which is not caused or subjectively justified by *any* intuition whatsoever.

It is easy to see how if *Descriptive Coherentism* is false, *Normative Coherentism* is fatally undermined. The problem for *Normative Coherentism* arises because it hangs (in a practical way) on the truth of *Descriptive Coherentism*. Remember that *Normative*

*Coherentism* is the thesis that coherentism is the way one ought to justify one's system of morality. Here we need only invoke a fairly uncontroversial point: "ought" implies "can." If intuitions are always the cause of reasoned judgments, we cannot engage in the process that coherentism requires of us. How can one then say that we ought to follow coherentism if we want to justify our system of morality?

Of course, a system of beliefs which is bounded by intransigent intuitions is not necessarily an incoherent system. As long as a system's intransigent intuitions are coherent, in fact, it seems inevitable that the rest of the system will be as well. Whether they are or not is a matter of empirical fact, but this fact is not in principle hidden from future research. If this were the case, both *Descriptive* and *Normative Coherentism* would be true. Moreover, even if both theses are (at present) false, there is nothing preventing them from becoming true in the future. The right set of cultural and sociological factors might produce a coherent set of intransigent intuitions in at least one person, and that is enough for *Descriptive Coherentism*. Such a person would then (automatically) be capable of meeting the "ought" requirement of *Normative Coherentism.* It would be a strange result if one's epistemic justification relied on factors which appear to be outside of one's rational control, but this is indeed where we find ourselves if we still want to argue that both theses are true.

As you might be able to see, there are two possible ramifications of the problem of intransigence. First, we may think that, because our ability to live up to *Descriptive Coherentism* is completely out of our rational control, *Normative Coherentism* should be abandoned and we should look elsewhere for a viable method of justification. Second, it

may be that *Normative Coherentism* is indeed true, in spite of the fact that we can never live up to it. If this is the case, it appears that we cannot ever obtain justification for our systems of morality, unless, as mentioned before, the right set of cultural and sociological factors aligns for us to be (miraculously!) justified.

At the very least, we should see the problem of intransigence resulting in a form of the MCS objection. This is because if we are not able to alter the judgments we identify with except when our intransigent intuitions are altered, it seems obvious that (at a minimum) no agreement among people with sufficiently different backgrounds is ultimately possible. Of course, there are again two possibilities here: only one coherent system is possible with the right set of cultural and sociological factors, or many coherent systems are possible with a number of different cultural and sociological factors. This is once more an empirical question. If the former is the case, there is still some question as to whether someone who (more or less) accidentally adopts such a system is really *justified* in holding that system.[12] If the latter is the case, this is a straightforward MCS objection.

Yet this is not the only problem for *Normative Coherentism.* The problem of intransigence relies on the perhaps controversial argument that all reasoned judgments are the causal product of intransigent intuitions. Though I think there is some good evidence for that position, the second problem arises even if we assume, *contra* all the arguments in the previous section, that some reasoned judgments *are* ultimately

[12] My intuitions on this are not clear one way or the other. Obviously such a person would believe for the right reasons (they hold a coherent system), but would they believe in the right *way*?

reasoned judgments. This would imply that *Descriptive Coherentism* is true and preempt the "ought implies can" argument against *Normative Coherentism*.

This next problem I have in mind is also a species of the MCS objection, and comes about much more straightforwardly. The main line of reasoning is this: because we have good evidence to presume that any given system will have at least *some* intransigent intuitions, it is overwhelmingly likely that multiple coherent systems of morality will arise. This is due to the nature of these intuitions; intransigent intuitions, I contend, will ultimately skew any process of reflective equilibrium in which they play a part. Thus when two different individuals have two incommensurable intransigent intuitions, they will almost always end up holding divergent coherent systems of morality. This result is even more likely if the two people in question have many incommensurable intransigent intuitions. I will call this problem the "limited problem of intransigence," and contend that it should be seen as a species of the MCS objection.

Let us return to an example from Haidt & Hersh (2001). Suppose you are a social conservative in this experiment who judges that consensual gay sex is wrong. What theoretical considerations might lead you to reject this judgment? Now suppose (for the sake of argument) that you (a conservative) also hold a belief akin to the harm principle: an action is only wrong when it harms someone non-consensually. Here we have a contradiction in your system of morality, for consensual gay sex is an ostensibly harmless act. But if the judgment you identify with is based on an intransigent intuition against gay sex, (as many of the subjects' judgments in Haidt's experiments are), noticing the contradiction is far more likely to make you reject the harm principle than it

is likely to make you override your original condemnative judgment of homosexual sex. Compare: if you are a (non-overriding) liberal and you lack a negative affective response to homosexual sex, and thus are not moved to condemn it, you will likely see no reason to abandon the harm principle at all. Given two individuals with different intransigent intuitions, contradictory systems of morality will most certainly arise through a process of reflective equilibrium. Indeed, it is not even necessary for *two* individuals to have intransigent intuitions; it is enough that one does and another does not (as in our example). This means that the two people in our example will be *justified* in holding their systems of beliefs, and this is what makes this an MCS objection. Insofar as the coherentist moral realist wants to provide an account of justification which produces just one set of justified beliefs, the limited problem of intransigence is nearly as damning as its less limited cousin.

Of course, the coherentist might still insist that at least one system has no intransigent intuitions at all, or (even) that there is some ideal system which includes a few intransigent intuitions which are correct *despite* their intransigence. This is a real possibility, but based on the evidence for the SI model adduced above, I think it is doubtful. We have no evidence that any individual lacks intransigent intuitions, and we certainly have no reason to believe that any given intransigent intuition must be correct. Here we must only keep the possibility in mind.

I contend that both the strong and weak types of the moral MCS objection are more telling than the MCS objection mentioned in section two. This is because of the differences between moral and non-moral systems of belief. The traditional version of

the MCS objection was defeated in section two by asking after the explanation that an isolated system would have for its coherence. (An isolated system, remember, is a coherent system of beliefs which rejects all experiential beliefs). We found that isolated systems cannot have explanations which are as good as non-isolated systems. A potential strategy for dealing with either of these MCS objections might then be to show that one or another *moral* system is isolated in a similar manner. Unfortunately, this approach will not work.

One minor problem is that it is not clear what is analogous to "experiential beliefs" in the moral case. We might solve this by stipulating that intuitions are the moral equivalent to experiential beliefs. But as soon as we see intuitions this way, we immediately come to notice a much larger problem: neither of the two coherent systems we are analyzing is an isolated system in this way. The liberal accepts some intuitions (notably intuitions arising out of the ethics of autonomy). The conservative accepts far more intuitions (from all three codes). No one rejects all intuitions, and even if someone did, this would only disqualify such a person's system, not the liberal's or the conservative's.

It appears as though what we would need for such a strategy is a means of determining which family of intuitions is a legitimate source of reasoned moral judgments, and which is not. We would then call isolated those systems which accepted only illegitimate sources of reasoned moral judgments. But then this strategy will surely not work, because that is precisely the reason we brought up the analogy with isolated

systems in the first place! This approach cannot solve the problem without begging the question.

There is a second approach we might take, but I am equally pessimistic about it. We might drop the attempt to draw an analogy to isolated systems in the non-moral case and directly compare the liberal and conservative systems on the grounds of how well each explains its own coherence. On those grounds, we might say that the liberal uses fewer argument patterns because he does not need to appeal to more than one family of intuitions in his explanation of his judgment. But using fewer families of intuitions is not obviously a virtue of a moral system. The conservative could easily point out that the liberal is merely excluding what to him (the conservative) are important intuitions. In the end, I do not think we can get an adequate handle on how well each system explains its own coherence, so this will not provide an adequate means of overcoming either of these moral versions of the MCS objection.

### 3.4 Implications for Coherentist Moral Realism

If we really cannot overcome these MCS objections through an explanatory requirement on coherent systems, it appears as though we have no means of calling one system less justified than another. Given the evidence that these systems will arise (through either the weak or strong version of this objection), we should turn our attention to how this fact affects the coherentist moral realist. Here it is important to emphasize that the view sees coherentism as a method for deriving objective moral *facts* (Brink 1989, 1989). Moreover, Brink's coherentist account in particular relies on realist second-order factual beliefs about one's system of moral beliefs. These beliefs are

realist because they ascribe the status of objectivity and fact to one's first-order moral beliefs. The problem that these versions of the MCS objection raise is that these second-order beliefs are most certainly rendered false (or at least severely undermined) by the knowledge that other people are equally justified in holding the coherent systems of belief they do. In short, coherentist realism requires that only one system of beliefs be coherent and thus justified, or those second-order realist beliefs on which it relies look incredibly implausible.

Consider why this is so. Imagine two moral realists (Martha and Christine) with coherent but contradictory systems of morality. Martha believes it is a fact of the matter that eating my dog Herbert is morally wrong. Christine believes that it is a fact of the matter that it is morally permissible. Each holds the second-order belief that she has *discovered* a moral fact about eating Herbert in virtue of the fact's overall coherence with her system of morality. Now suppose that Martha knows about the existence of Christine (and vice versa) and that each knows that the other's system of belief is just as coherent as hers, but incorporates a belief which contradicts hers. Because Martha thinks that her moral belief is a *fact of the matter*, she must believe that Christine is wrong, and Christine must believe the same thing about Martha's belief. Now it is easy to ask: who is right? Both are equally justified, so there seems to be no basis for concluding that either one or the other is right. This means that for the question over which their answers diverge (the permissibility of Herbert-eating), there can be *no one realist answer.*

Such a possibility shows one of two things: either (1) one of the aforementioned realists is wrong, and there is genuinely no way to tell which one it is (in which case moral facts are sometimes genuinely indeterminate), or (2) moral objectivism is false (in which case we retain determinacy of moral facts by relativizing those facts to a particular coherent systems of morality).

Though he thinks that these cases are far more rare than I do, Brink favors the first answer. He argues that in these types of cases, one side is simply *systematically mistaken*. There is a fact of the matter as to who is correct, but one side is simply wrong about almost everything, and the other right about almost everything. He does, however, note that the more that one claims that one's opponent is simply systematically mistaken, the less justified one's system of belief appears to be. This is certainly the case when matters of disagreement are between large and more or less comprehensive ideologies like liberalism and conservatism; one cannot merely hand-wave these disputes away by saying that either liberals or conservatives are systematically mistaken (Brink 1989, 200).

Brink also notes that for some moral disputes, the realist might accept that "there are no uniquely correct answers" (Brink 1989, 202). He admits that this strategy must be confined to only a few cases, because if a realist were forced to be agnostic about *most* moral facts, he would hold a very strange variant of realism (ibid.).

I do not think either of these strategies is adequate. First, it is not clear that admitting even one case of systematic mistakenness is acceptable to the moral realist. Calling a position "systematically mistaken" seems to be terminologically begging the

question. In order to be *systematically* mistaken, one must hold a view so different from another view that no amount of reasoned dialogue could possibly bring about convergence. In these cases, calling one side mistaken cannot add up to more than an ad hominem. After all, the whole problem is that it is impossible to tell (from an objective standpoint) exactly who is mistaken!

Second, Brink appears to think that as long as the *number* of irresolvable moral disputes is small, the fact that there are no uniquely correct answers is acceptable for the realist. Though I believe we have evidence sufficient to believe that moral disagreements are large in quantity, this surely cannot be the deciding factor. Rather, some moral matters are more important than others; if the permissibility of stealing candy from convenience stores has no uniquely determined answer, this might be acceptable. But based on the cases of moral dumbfounding that Haidt reports, irresolvable differences may in fact be over questions far more pressing than that. We should see that in accepting coherentist moral realism we are in danger of licensing relativism over very basic questions of value: certainly the permissibility of "deviant" sexual practices, perhaps even the permissibility of access to abortion, the permissibility of certain wars, the structure of taxation, etc. Indeed, much of what has been popularly labeled the "culture wars" in the United States may come down to fundamentally irresolvable and intransigent intuitions. Justified relativism in these areas has *dramatic* social and political ramifications, and cannot simply be brushed off by the coherentist realist as an acceptable price to pay for realism. If coherentist moral realism is to remain

a viable contender in the field of moral epistemology, it must come up with resources to deal with these two MCS objections.

### 3.5 Section Conclusion

As we have seen, there are good reasons to think that moral judgments cause emotions, and that the moral judgments with which we identify are always caused by intuitive moral judgments. This possibility leads to two serious problems, both of which constitute MCS objections to coherentist moral realism.

It is important to note exactly what this critique is committed to. First, neither the problem of intransigence nor the limited problem of intransigence is a claim about the necessary relations between intuition and reasoned judgment. The claim in the problem of intransigence is not that reasoned judgment is *necessarily* caused by intuition, just that as a matter of empirical fact, it *always is*. And similarly, the limited problem of intransigence is not that all moral systems *necessarily* have intransigent intuitions, but that as a matter of empirical fact, they all do. As such, both could be argued against by denying the empirical claims on which they rest. The former could easily be falsified by pointing to an example of a reasoned judgment which is not caused by an intuition, and the latter could be falsified by showing that at least one coherent moral system does not contain any intransigent intuitions at all. Second, it is important to emphasize that these two problems do not rely on traditional claims in metaethics; neither claims that moral judgments *are* emotions, or that moral judgments are "always motivating." It rather takes a more limited approach, claiming that moral judgments cause emotions, but that intuitive moral judgments are always causally primary (or in the

case of the limited problem, primary *often enough*). It thus avoids the non-cognitivist quagmire into which so much twentieth century metaethical debate has sunk, and represents an empirically-based cognitivist critique of coherentist moral realism.

## 4. CONCLUSION

In this thesis I have argued that we have good reasons to prefer a coherentist moral realist account of justification in moral epistemology. Such an account is not committed to constructivism in ethics, avoids the problems of foundationalism, and can be modified so as to successfully resist three of the most important traditional objections to coherentism. Even so, I pointed to two types of MCS objection which accept that moral judgments have all the features a moral realist expects them to, but appears to cause a real problem for the view nonetheless.

I remain hopeful that we may find a way out of the two versions of the MCS objection that I proposed in section three. But I also think that any successful version of coherentist moral realism must meet two important obstacles brought up by these objections. First, it must explain how moral reasoning can improve the coherence of a person's system of belief when all of the reasoned judgments in such a system are caused by intuitions. Second, it must explain how we can derive a single coherent moral system which is more justified than all the alternatives. Or, failing that, it must at least identify a set of equally justified coherent systems which do not differ in their answers to our most serious moral questions. In my later work, deriving just such a system (or set of systems) will be my goal.

## REFERENCES

Audi, Robert. "Moral Epistemology and the Supervenience of Ethical Concepts." *The Southern Journal of Philosophy*. 49.S1 (1990): 1-24. Print.

Blackburn, Simon. "How to Be an Ethical Antirealist." *Midwest Studies in Philosophy* 7.1 (1988): 361-375. Print.

BonJour, Laurence. *The Structure of Empirical Knowledge.* Cambridge: Harvard University Press, 1985.

Brink, David Owen. *Moral Realism and the Foundations of Ethics.* New York: Cambridge University Press, 1989.

—. "Rawlsian Constructivism in Moral Theory." *Canadian Journal of Philosophy* 17.1 (1987): 71-90. Print.

Haidt, J., F. Bjorklund and S. Murphy. "Moral Dumbfoujnding: When Intuition Finds no Reason." *Unpublished Manuscript.* University of Virginia. Online access: http://people.virginia.edu/~jdh6n/publications.html (2000).

Haidt, J., S. H. Koller and M. G. Dias. "Affect, Culture, and Morality or Is it Wrong to Eat Your Dead Dog?" *Journal of Personality and Social Psychology* 65.4 (1993): 613-628. Print.

Haidt, Jonathan & Hersh, Matthew A. "Sexual Morality: The Emotions of Conservatives and Liberals." *Journal of Applied Psychology* 31.1 (2001): 191-221. Print.

Haidt, Jonathan. "The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review* 108.4 (2001): 814-834. Print.

Harman, Gilbert. "Moral Explanations of Natural Facts: Can Moral Claims Be Tested Against Reality?" *Southern Journal of Philosophy* 24.S1 (1986): 57-68. Print.

Hempel, Karl. "Laws and Their Role in Scientific Explanation." Boyd, Richard, Philip Gasper and J. D. Trout. *The Philosophy of Science.* Cambridge: MIT Press, 1991. 299-316.

Kenny, Anthony. *Action, Emotion, and Will.* New York: Routledge, 2003.

Kitcher, Philip. "Explanatory Unification." Boyd, Richard, Philip Gasper and J. D. Trout. *The Philosophy of Science.* Cambridge: MIT Press, 1991. 329-348.

Lazarus, Richard. *Emotion and Adaptation.* New York: Oxford University Press, 1991.

—. "Relational Meaning and Discrete Emotions. " Scherer, Klaus R, Angela Schorr, & Tom Johnstone. *Appraisal Processes in Emotion.* New York, NY: Oxford University Press, 2001.

Lehrer, Keith. *Theory of Knowledge.* Boulder: Westview Press, 2000.

Leventhal, H. and K. Scherer. "The Relationship of Emotion to Cognition: A Functional Approach to a Semantic Controversy." *Cognition and Emotion* 1.1 (1987): 3-28. Print.

Nussbaum, Martha C. *Upheavels of Thought: The Intelligence of Emotions.* New York: Cambridge University Press, 2001.

Prinz, Jesse. *The Emotional Construction of Morals.* New York: Oxford University Press, 2007.

Rawls, John. *A Theory of Justice.* 1971. Reprinted. Cambridge: Harvard University Press, 1999.

Ross, W. D. *The Right and the Good.* 1930. Reprinted. New York: Oxford University Press, 2002.

Rozin, P., et al. "The CAD Triad Hypothesis: A Mapping Between Three Moral Emotions (Contempt, Anger, Disgust) and Three Moral Codes (Community, Autonomy, Divinity)." *Journal of Personality and Social Psychology* 76.4 (1999): 574-586. Print.

Scarantino, Andrea. "Insights and Blindspots of the Cognitivist Theory of Emotions." *British Journal for the Philosophy of Science* 61.4 (2010): 729-768. Print.

Teasdale, J. D. "Multi-level Theories of Cognition-Emotion Relations." Dalgleish, T. and M. J. Power. *Handbook of Cognition and Emotion.* Chichester: John Wiley and Sons, 1999. 665-681.

van Reekum, C and K. R. Scherer. "Levels of Processing in Emotion-antecedent Appraisal." Matthews, G. *Cognitive Science Perspectives on Personality and Emotion.* Amsterdam: Elsevier, 1997. 259-300.

Zajonc, R. B. "Feeling and Thinking: Preferences Need No Inference." *American Psychologist* 35.2 (1980): 151-175. Print.

# VITA


Name: Ross T. Colebrook

Address: c/o Gary Varner, 314 Bolton Hall, Texas A&M University, College Station, TX 77843-4237

Email Address: colebrooked@gmail.com


Education: B.A., Political Science & Philosophy, American University 2009
M.A., Philosophy, Texas A&M University, 2011